DOCUMENT RESUME

ED 056 560 FL 002 480

TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications.
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-25-26-71
PUB DATE	71
NOTE	151p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Acoustic Phonetics; Anatomy; Articulation (Speech); *Auditory Perception; Consonants; Data Analysis; Ears; *Electronic Data Processing; *Experiments; *Language Research; Memory; Neurological Organization; Phonemes; Phonetics; *Phonology; Physiology; Reading; Spectrograms; Syllables; Vowels

ABSTRACT

This collection on speech research presents a number of reports of experiments conducted on neurological, physiological, and phonological questions, using electronic equipment for analysis. The neurological experiments cover auditory and phonetic processes in speech perception, auditory storage, ear asymmetry in dichotic listening, auditory perception and processing, cerebral reaction to speech signals, and the relationship between speech and reading disability. The physiological experiments examine articulatory muscles, control of tongue position, laryngeal muscles, velopharyngeal mechanism, and vocal tract size normalization. The phonological investigations consider tense-lax features in some English vowels and "Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English." One report covers the latest development in the Haskins electronic system for collecting and processing data in research. (VM)

ED056560

SR-25/26 (1971)

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 January - 30 June 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

FL002 480

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use.)

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

Information Systems Branch, Office of Naval Research
Contract N00014-67-A-0129-0001
Req. No. NR 048-225

National Institute of Dental Research
Grant DE-01774

National Institute of Child Health and Human Development
Grant HD-01994

Research and Development Division of the Prosthetic and Sensory Aids Service, Veteran Administration
Contract V-1005M-1253

National Institutes of Health
General Research Support Grant FR-5596

Connecticut Research Commission
Grant Award No. RSA-70-9

CONTENTS

I. Manuscripts and Extended Reports

Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study -- M. Studdert-Kennedy, D. Shankweiler, and D. Pisoni..... 1

An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage -- C.J. Darwin, M.T. Turvey, and R.G. Crowder..... 13

Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening -- C.J. Darwin..... 29

The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables -- R.J. Porter..... 35

Auditory Evoked Potentials During Speech Perception -- C.C. Wood, W.R. Goff, and R.S. Day..... 41

Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability -- I.Y. Liberman..... 51

The EMG Data System -- D.K. Port..... 67

Electromyography of the Articulatory Muscles: Current Instrumentation and Technique -- H. Hirose..... 73

Action of the Extrinsic Musculature in the Control of Tongue Position: Preliminary Report -- K.S. Harris..... 87

Electromyography of the Intrinsic Laryngeal Muscles During Phonation -- T. Gay, H. Hirose, M. Strome, and M. Sawashima..... 97

An Electromyographic Study of Laryngeal Adjustments During Speech Articulation: A Preliminary Report -- H. Hirose..... 107

The Velopharyngeal Mechanism: An Electromyographic Study A Preliminary Report -- F.B. Berti..... 117

An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels -- L.J. Raphael..... 131

Vocal Tract Size Normalization in the Perception of Stop Consonants -- T.C. Rand..... 141

ABSTRACT: Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English -- L.J. Raphael..... 147

II. Publications and Reports..... 149

Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study*

M. Studdert-Kennedy,[+] D. Shankweiler,[++] and D. Pisoni[+++]
Haskins Laboratories, New Haven

ABSTRACT

The distinction between auditory and phonetic processes in speech perception was used in the design and analysis of an experiment. Earlier studies had shown that dichotically presented stop consonants are more often identified correctly when they share place of production (e.g., /ba-pa/) or voicing (e.g.,/ba-da/) than when neither feature is shared (e.g.,/ba-ta/). The present experiment was intended to determine whether the effect has an auditory or a phonetic basis. Increments in performance due to feature-sharing were compared for synthetic stop-vowel syllables in which formant transitions were the sole cues to place of production under two experimental conditions: (1) when the vowel was the same for both syllables in a dichotic pair, as in our earlier studies, and (2) when the vowels differed. Since the increment in performance due to sharing place was not diminished when vowels differed (i.e., when formant transitions did not coincide), it was concluded that the effect has a phonetic rather than an auditory basis. Right-ear advantages were also measured and were found to interact with both place of production and vowel conditions. This outcome suggests that inhibition of the ipsilateral signal in the perception of dichotically presented speech occurs during or immediately before phonetic analysis.

Current accounts of speech perception emphasize process and divide the process into a hierarchy of stages: auditory, phonetic, phonological, and so on (see, for example, Fry, 1956; Chistovich et al., 1968; Studdert-Kennedy, in press). The distinction between phonetic and higher levels is commonly accepted in linguistic theory and is readily demonstrated in behavior. But the distinction between auditory and phonetic levels is less easily demonstrated and is not widely recognized. The auditory stage (or stages) refers to transformation of the acoustic waveform into a set of time-varying psychological dimensions (pitch, loudness, timbre, duration) roughly corresponding to dimensions measurable in a spectrogram. The phonetic stage refers to transformation of psychological (auditory) dimensions into phonetic features. We

---

* The results of this study were reported at the 81st Meeting of the Acoustical Society of America, Washington, D.C., April 1971.

+ Also Graduate Center and Queens College, City University of New York.

++ Also University of Connecticut.

+++ Now at Indiana University.

have argued elsewhere (Studdert-Kennedy and Shankweiler, 1970) that, while the auditory transformation may be accomplished by the general auditory system common to both cerebral hemispheres, the phonetic transformation is accomplished largely, if not exclusively, by specialized mechanisms in the language-dominant hemisphere.

We will not repeat the argument here. But among the reasons for positing a single phonetic processing system is an interaction between left- and right-ear inputs repeatedly observed in dichotic experiments: the initial stop consonants of dichotically presented CV (or CVC) syllables, differing only in those stops, are more accurately identified if the two segments have a phonetic feature in common, (Shankweiler and Studdert-Kennedy, 1967). Figure 1 (based on Table IV of Studdert-Kennedy and Shankweiler, 1970) displays the effect. The probability that both initial stops will be correctly identified is greater if the two segments have the same value on the phonetic features of place (e.g., /ba-pa/) or voicing (e.g., /ba-da/) than if they have neither feature in common (e.g., /ba-ta/).

We have interpreted this interaction as evidence that dichotic speech inputs converge on a single cerebral center before the extraction of phonetic features. We suggested, further, that "duplication of the auditory information conveying the shared feature value gives rise to the observed advantage" (Studdert-Kennedy and Shankweiler, 1970, p.589). However, there are at least two stages at which the advantage might arise: (1) during extraction of phonetic features from the auditory transforms (the interpretation quoted above); or (2) during output of a response from the phonetic system.[1] The first interpretation attributes the advantage to shared characteristics of the inputs (signals) to the phonetic system: phonetic analysis of the two sets of auditory parameters is facilitated if the two sets have certain auditory features in common. The second interpretation attributes the advantage to shared characteristics of the outputs (messages): correct responses from the phonetic component are facilitated by shared phonetic features rather than by shared auditory features.

The present experiment was designed mainly to distinguish between these two interpretations. We may clarify the argument by considering the set of syllables used. Table I lists four stop consonants (/b,p,d,t/) and their possible combinations into dichotic pairs. Note that there are two pairs sharing place (/b-p/, /d-t/), two sharing voice (/b-d/, /p-t/) and two sharing neither feature (/b-t/, /d-p/). We are most interested in the two pairs sharing place, since it is these that permit us to compare the effects of auditory and phonetic commonalty.

Figure 2 illustrates the comparison. The figure displays stylized spectrographic patterns of the eight synthetic CV syllables used in this study. They are formed from all possible combinations of the four stop consonants (/b,p,d,t/) with two vowels (/i,u/). No release burst was included in the synthesis so that all information concerning place of articulation is conveyed by the second- and, to some extent, third-formant transitions. All within-

[1]The advantage might also arise during extraction of auditory information from the acoustic signal. The effect would then be due to a relatively low, perhaps subcortical, level of the perceptual process. As will be seen, this possibility, though difficult to test directly, was ruled out by implication from the results of the experiment.

The Percentage of Trials on Which Both Responses Were Correct as a Function of the Consonantal Feature Shared by Dichotic CV Pairs

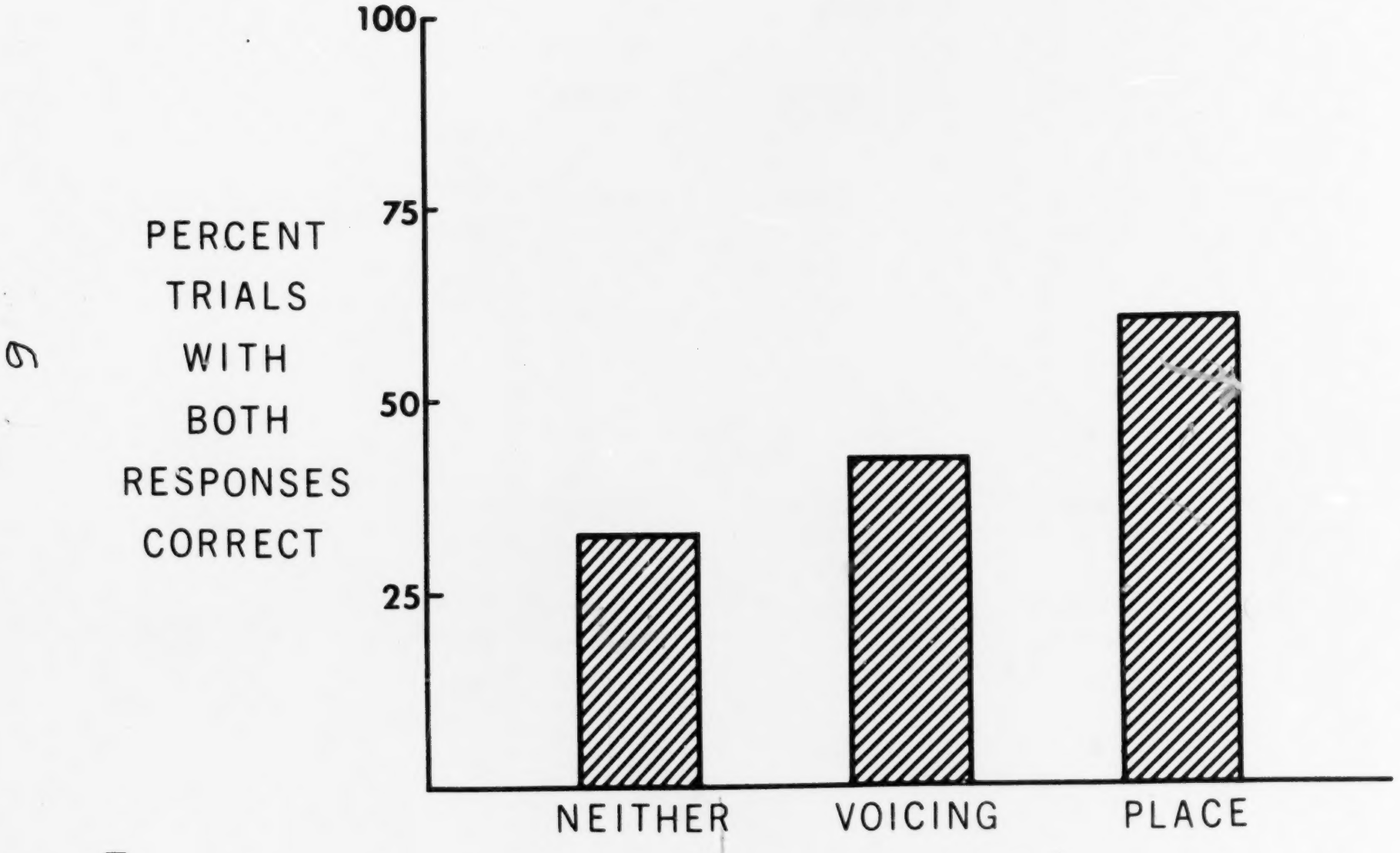
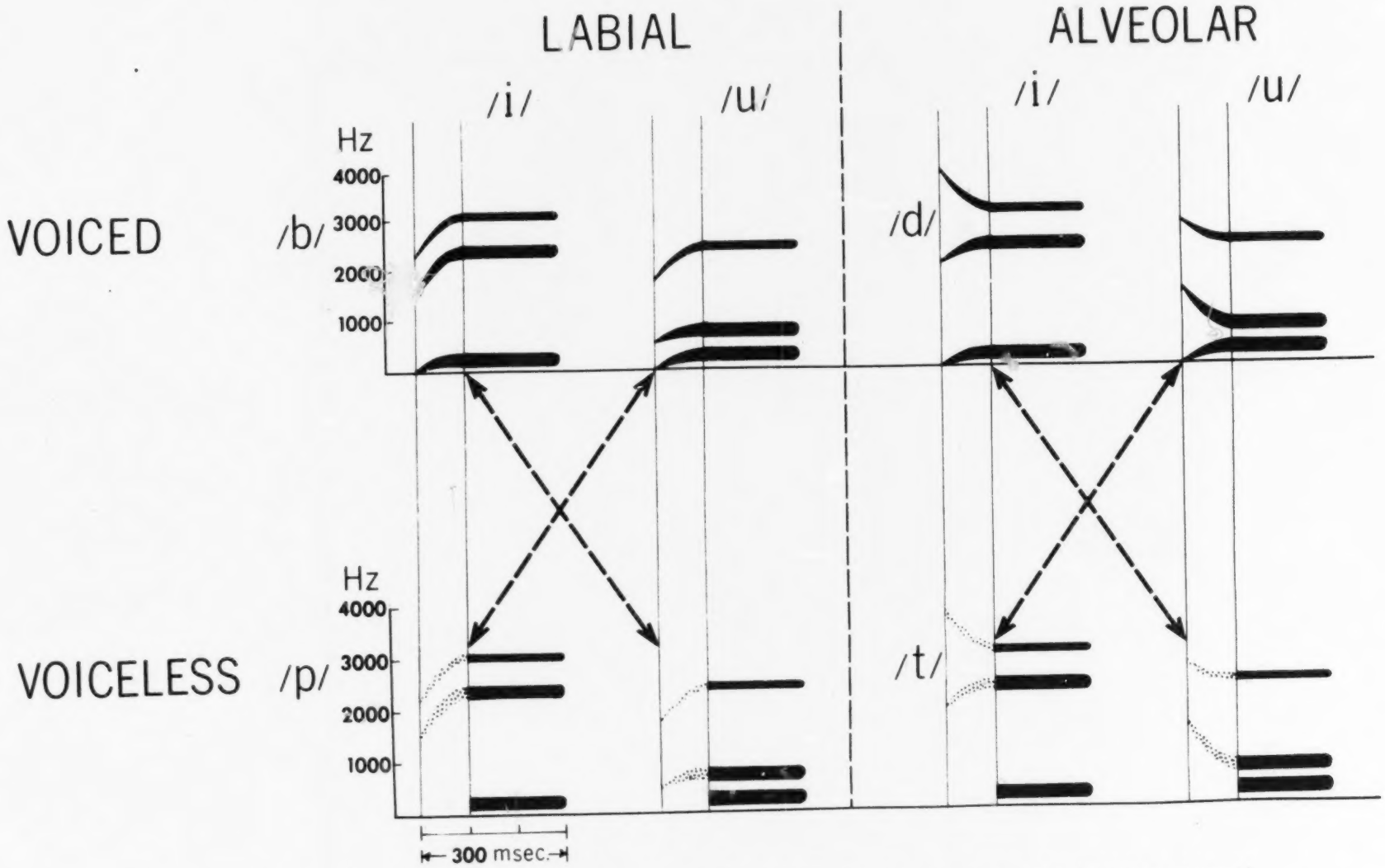


Fig. 1

Schematic Spectrograms of Eight Synthetic CV Syllables



Syllables within a column share consonantal place of articulation and vowel.

Syllables across a diagonal share place of articulation but not vowel.

Fig. 2

column pairs share both place of consonantal articulation and following vowel: they therefore have identical formant transitions. Cross-column pairs (/bi-pu, bu-pi, di-tu, du-ti/) share place of consonantal articulation but not following vowel: they therefore have different formant transitions. In other words, within-column (same vowel) pairs share both phonetic and auditory information; cross-column (different vowel) pairs shared only phonetic information. We may now compare performance on these two types of dichotic pair. If the advantage due to sharing a feature has an auditory basis, we would expect the advantage to be greater for place-sharing dichotic pairs that also share the same vowel than for corresponding pairs that have different vowels. On the other hand, if the advantage has a phonetic basis, we would expect no difference in performance on these same pairs between the two experimental conditions of "vowel same" and "vowel different."

TABLE I

Paired Combinations of Four Stop Consonants According to Features of Voicing and Place of Articulation

	Place of Articulation	
	Labial	Alveolar
Voiced	b	d
Unvoiced	p	t

Pairs Sharing:

Place Alone	Voicing Alone	Neither Feature
b-p	b-d	b-t
d-t	p-t	d-p

Finally, a subsidiary purpose of the present experiment was to determine the effect of auditory commonalty or contrast on the right-ear advantage typically observed for stop consonants in dichotic studies. We defer elaboration of this matter to the discussion.

METHOD

The eight three-formant CV syllables were synthesized on the Haskins Laboratories parallel resonant synthesizer. Each syllable had a duration of 300 msecs : formant transitions lasted 40 msec, steady state portions, 260 msecs. For the voiced consonants all three formants began at the same instant; for the voiceless consonants the first formant was cut back by 70 msec, and the upper formants were aspirated over this period. The pitch contour of each syllable fell linearly from 130 Hz to 90 Hz.

Two dichotic tapes were prepared by a computer-controlled procedure that permits precise alignment of syllable onsets. Voiced/voiceless pairs (i.e., those sharing place: /b-p/, /d-t/) were aligned so that the aspirated formants of one syllable began at the same instant as the voiced formants of the other. On one tape, the vowels of any dichotic pair were the same (either /i/ or /u/); on the other tape, the two vowels were different. There are twelve possible ordered pairs of syllables contrasting in their initial consonants (ordering refers to channel orientation). Each pair occurred ten times in a randomized test order, with the restriction that each pair occurred five times in the first sixty trials, five times in the second sixty trials.

Sixteen university students volunteered as subjects and were paid for their work. All were right-handed native speakers of English and had no known hearing loss. They were run as four groups over two days in a balanced design, distributing all order effects equally over the two experimental conditions. On a given day, the subjects began with an eighty-item monaural identification tape, forty items to the left ear, forty to the right. They then took a twenty-four-item practice dichotic tape. Finally, they took the assigned test tape twice reversing earphones after the first run to distribute channel effects equally over the ears. For the dichotic test they were told that the two consonants on any trial would always be different; they were instructed to identify both of them, drawing from the set /b,p,d,t/, to write their answers on a sheet, and to give their more confident response first.

One subject scored less than 90 percent on the monaural identification test and displayed a strong left-ear advantage on every data analysis. He was omitted from the group analysis, reducing the total number of trials to 1,800, 120 from each of the fifteen subjects.

RESULTS

Figure 3 displays the main results. For both experimental conditions (vowel same, vowel different), the percentage of trials having both responses correct is greater for those dichotic pairs that have a feature in common. The effect is significant by analysis of variance ($p < .001$). In previous studies (cf., Figure 1) more advantage accrued to pairs sharing place than to pairs sharing voicing. Here, there is no significant difference between the two classes of dichotic pair: subjects varied in whether they gave their highest performance on place-sharing or voice-sharing pairs, so that there was significant subject-by-feature-shared interaction ($p < .001$). No subject gave his highest performance on pairs having no feature in common.

Turning to the result of most interest for the present study, we note that there is no significant effect of the following vowel. The slight advantage for place-sharing pairs that precede different vowels was present for both labial (5 percent) and alveolar (6 percent) pairs but was not significant.

Finally, we consider the ear advantages. Table II displays the distribution of correct responses over the ears for trials on which only one response was correct (the only trials on which an ear advantage has an opportunity to occur).[2] The columns headed (R-L/R+L) 100 provide a measure of the ear advantage: the index ranges from 0 to ± 100 with negative values indicating a

[2]The greater number of such trials when neither feature was shared is entailed by the smaller number of both correct trials under that condition.

TABLE II

Distribution of Correct Responses Over Ears for One-Correct Trials Only

Feature Shared	Vowel Same			Vowel Different			Total		
	R	L	$\frac{(R-L)}{(R+L)}100$	R	L	$\frac{(R-L)}{(R+L)}100$	R	L	$\frac{(R-L)}{(R+L)}100$
Place	326	208	22	281	193	19	607	401	20
Voicing	289	236	10	279	255	5	568	491	7
Neither	439	329	14	421	331	12	860	660	13
Total	1054	773	15	981	779	11	2035	1552	13

R = Number correct on right ear

L = Number correct on left ear

The Percentage of Trials on Which Both Responses Were Correct as a Function of the Consonantal Feature Shared by Dichotic CV Pairs Under Two Vowel Conditions

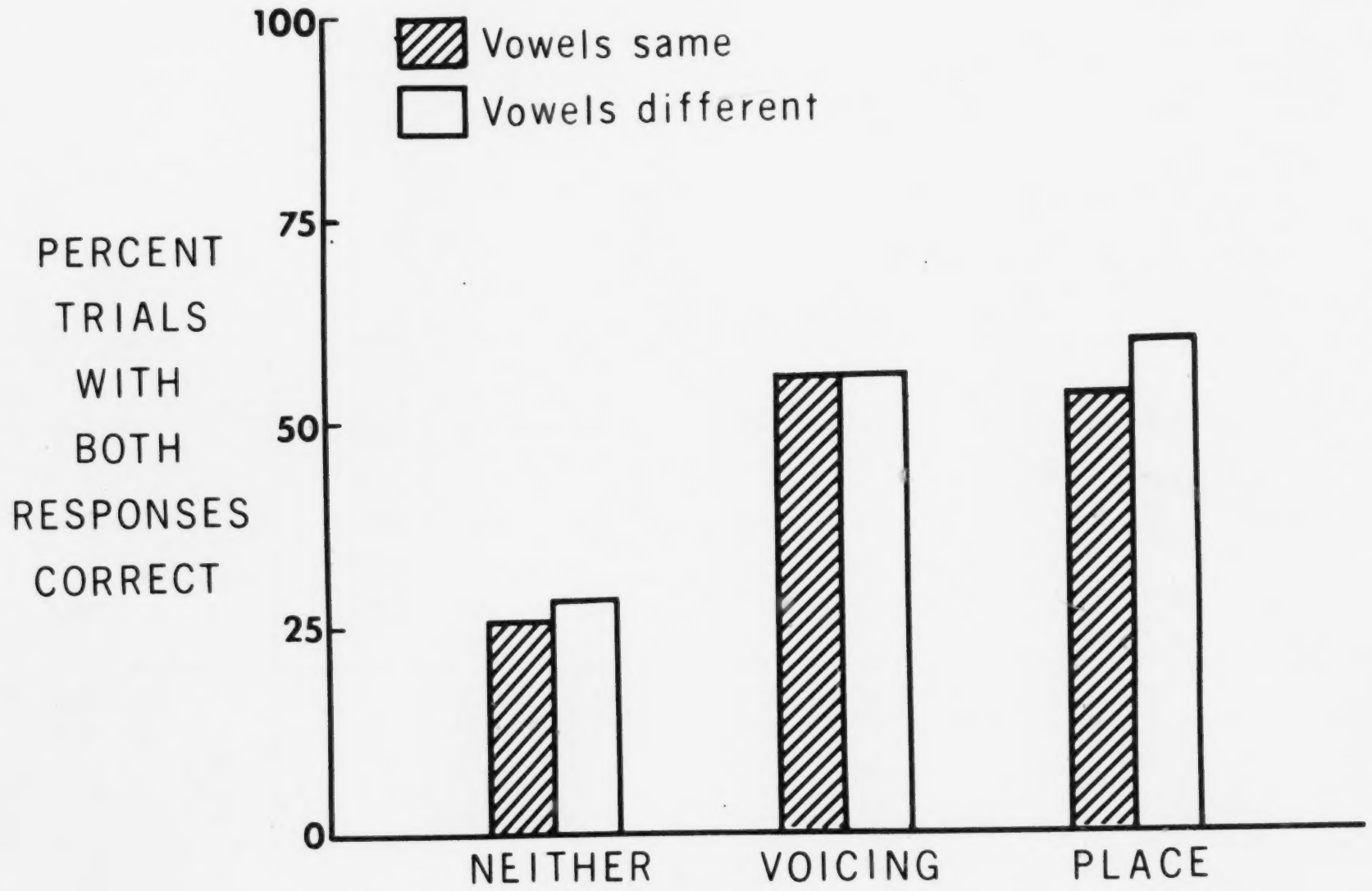


Fig. 3

left-ear advantage, positive values a right-ear advantage. All indices are positive and the ear effect is highly significant by analysis of variance ($p < .001$); its variation across feature conditions falls short of significance at the .05 level. There is no reliable difference in the ear effects for the two vowel conditions: the tendency toward a larger laterality index when vowels are the same than when they are different is not significant.

However, analysis of the one-correct data into separate place values reveals complexities: there is significant, three-way interaction between ears, vowel condition, and place value ($p < .05$). Table III shows that for alveolar pairs the laterality index is greater when vowels are the same than when vowels are different; for labial pairs the reverse is true. We may note, further, that the alveolar ear-by-vowel interaction is primarily due to a drop in right-ear performance when vowels are different, while the labial ear-by-vowel interaction is largely due to a rise in left-ear performance when vowels are the same. Summing over vowel conditions, we note no significant difference in the laterality effect for the two place values.

DISCUSSION

The main outcome is predicted by the phonetic interpretation: the gain in performance for feature-sharing dichotic pairs arises from commonalty in the phonetic message rather than in the acoustic signal or its auditory transform. From this we may draw two inferences. First, a response is composed by integration of the outputs from distinct phonetic feature processors. Second, activation of a feature processor for one response facilitates its activation for another temporally contiguous response. The same statements might serve to describe a short-term response bias leading to errors of feature substitution in speaking of the kind described by Fromkin (1970). However, in the present instance, repetition of a feature in successive responses is not a random, internally generated error but the apt sequel of auditory information extracted from paired signals. The effect is therefore perceptual.

At the same time, the results justify the distinction between auditory and phonetic processes upon which the experiment was based, since commonalty at the two levels affects overall performance and the laterality effect differently. Phonetic feature-sharing facilitates performance but has little or no effect on the ear advantage. Auditory similarity or contrast affects the ear advantage (Table III) but not performance. We conclude that phonetic and auditory transformations are indeed distinct processes. Furthermore, the phonetic transformation seems to be accomplished by a single system to which both dichotic inputs have access.

We turn now to the ear advantages. We have argued elsewhere (Studdert-Kennedy and Shankweiler, 1970) that auditory-to-phonetic transformation may be the prerogative of the language-dominant cerebral hemisphere. At the same time, the minor hemisphere is evidently specialized for recognition of complex auditory patterns (Milner, 1962; Kimura, 1964, 1967; Shankweiler, 1966; Darwin, 1969). The interaction between ears, feature-value shared, and vowel condition in the present study (see Table III) may reflect, in part, this functional dissociation of the hemispheres. Study of Figure 2 will show that the most marked formant transition contrast is between alveolar pairs followed by different vowels (/di-tu, du-ti/). If we assume that auditory analysis of both inputs is attempted by both hemispheres, we might expect that these pairs, with their

TABLE III

Distribution of Correct Responses Over Ears
and Place Values for One-Correct Trials Only

Feature Value Shared	Vowel Same			Vowel Difference			Total		
	R	L	$\frac{(R-L)}{(R+L)}$ 100	R	L	$\frac{(R-L)}{(R+L)}$ 100	R	L	$\frac{(R-L)}{(R+L)}$ 100
Alveolar	161	87	30	122	93	13	283	180	21
Labial	165	121	15	159	100	23	324	221	19
Total	326	208	22	281	193	19	607	401	20

R = Number correct on right ear

L = Number correct on left ear

conflicting transitions, would present the greatest analytic problem and that this problem would be more difficult for the right-ear/left-hemisphere system than for the left-ear/right-hemisphere system.[3] The results bear this out: it is precisely these pairs that lower right-ear performance and contribute most strongly to the observed interaction.

We should be clear that we are here accounting not for a reversal of the ear advantage but for a reduction in its size due to lowered right-ear performance under one condition of this experiment. We should not confuse this reduction with the generally lower left-ear performance observed in dichotic speech studies. The latter may be attributed to loss of auditory information arising from interhemispheric transfer of the left-ear signal to the dominant hemisphere for phonetic processing (see Studdert-Kennedy and Shankweiler, 1970), while the reduced right-ear performance under one condition of this experiment is here attributed to increased interference of the left-ear signal with the right-ear signal during auditory analysis in the left hemisphere.

Clearly this account is not complete, since it leaves unexplained the rise in left-ear performance on labial consonants when vowels are the same. However, detailed explanation is of less importance than the fact of the interaction. The finding that the vowel condition affects stop consonant perception differently for the two ears (Table III) is the first evidence of central auditory interaction between dichotic speech inputs. From this we may infer that inhibition of the ipsilateral signal under dichotic stimulation (see Milner, Taylor, and Sperry, 1968) occurs not in the pathways to the cerebral hemispheres but after central auditory analysis, either at the auditory-phonetic interfacing or during phonetic analysis. In this regard, we may note that the laterality effect for speech is only obtained if both signals are perceived as speech: contralateral white noise (Shankweiler and Halwes, unpublished data), noise limited to the speech band (Darwin, 1971), or pure tones (Day and Cutting, 1970) do not produce an ear advantage. This, too, would seem to implicate phonetic rather than auditory analysis as the primary level of dichotic competition.

To sum up, this study has provided further grounds for distinguishing between auditory and phonetic levels of speech processing. The results suggest that both signals of a dichotically presented syllable pair are transmitted to a single phonetic processor and that correct output from that processor is facilitated if the two messages have phonetic features in common. At the same time, they suggest that inhibition of the ipsilateral signal in the perception of dichotically presented speech may occur during or immediately before phonetic analysis.

---

[3] Superiority of the right hemisphere in auditory pattern recognition has so far been shown only for nonspeech patterns. The possibility of left-hemisphere superiority in the analysis of patterns peculiar to speech (such as formant transitions), due to its possession of specialized auditory feature processors, cannot be excluded. This possibility is currently being investigated experimentally at Haskins Laboratories. In the present account, we are tentatively assuming, on the basis of the cited dichotic work with nonspeech patterns, that the left hemisphere is inferior in the resolution of conflicting ipsilateral and contralateral auditory patterns.

REFERENCES

Chistovich, L.A., A. Golinsina, V. Lublinskaja, T. Malivnikova, and M. Zukova, (1968) Psychological methods in speech perception research. Z. Phon. Sprachwiss. u. Komm. Fschg. 21, 102-106.

Darwin, C.J. (1971) Dichotic forward and backward masking of speech and nonspeech sounds. Paper presented at 81st Meeting of Acoustical Society of America, April 1971, Washington, D.C.

Darwin, C.J. (1969) Auditory perception and cerebral dominance. Unpublished Ph.D. thesis, University of Cambridge.

Day, R.S. and J.E. Cutting (1970) Perceptual competition between speech and nonspeech. J. Acoust. Soc. Amer. 49, 85(A). (Text in Haskins Laboratory Status Report, SR/24.)

Fromkin, V.A. (1970) Tips of the slung - or - to err is human. Univ. of California at Los Angeles: Working Papers in Phonetics 14, 40-79.

Fry, D.B. (1956) Perception and recognition in speech. In For Roman Jakobson, ed. by M. Halle, H.G. Lund, H. McClean, and C.H. van Schooneveld. The Hague: Mouton.

Kimura, D. (1964) Left-right differences in the perception of melodies. Quart. J. Exp. Psychol. 16, 355-358.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Milner, B. (1962) Laterality effects in audition. In Interhemispheric Relations and Cerebral Dominance, ed. by V.B. Mountcastle. Baltimore: John Hopkins Univ. Press.

Milner, B., L. Taylor, and R.W. Sperry. (1968) Lateralized suppression of dichotically-presented digits after commissural section in man. Science 161, 184-185.

Shankweiler, D. (1966) Effects of temporal-lobe damage on perception of dichotically presented melodies. J. Comp. Physiol. Psychol. 62, 115-119.

Shankweiler, D. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.

Studdert-Kennedy, M. (In press) The perception of speech. In Current Trends in Linguistics, Vol. XII, ed. by T.A. Sebeok The Hague: Mouton. (Also in Haskins Laboratories Status Report SR/23, 1971, pp. 15-48).

Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Am. 48, 579-594.

An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage*

Christopher J. Darwin,[+] Michael T. Turvey,[++] and Robert G. Crowder[+++]

ABSTRACT

Four experiments are reported on the partial report of material presented auditorily over three spatially different channels. When partial report was required by spatial location, it was superior to whole report if the cue came less than two (Exp. I) or four (Exp. II) seconds after the end of the stimuli. When partial report was required by semantic category (letters/digits), the relation between it and whole report depended on whether the subject was asked also to attribute each item to its correct spatial location. When location was required, partial report was lower than whole report and showed no significant decay with delay of the partial report indicator (Exp. III), but when location was not required, partial report was superior to whole report for indicator delays of less than two seconds (Exp. IV). This superiority was, however, much less than that found when partial report was required by spatial location. These results are compatible with a store which has a useful life of around two seconds and from which material may be retrieved more easily by spatial location than by semantic category.

INTRODUCTION

The concept of brief sensory storage has played a central role in recent discourse on the nature of human information processing (e.g., Neisser, 1967; Haber, 1969; Hunt, 1971). The proposition is that sensory data is initially represented in a literal, labile form for a brief duration during the course of conversion into a relatively more persistent, categorized form.

The sensory store which has received the most attention, and which consequently we know most about, is in vision. The characteristics of that store, called iconic by Neisser (1967), have been isolated via the delayed partial-sampling procedure of Sperling (1960) and Averbach and Coriell (1961).

---

*Paper to appear in Cognitive Psychology.

+Haskins Laboratories, New Haven; now at the University of Sussex.

++Haskins Laboratories, New Haven, and the University of Connecticut, Storrs.

+++Yale University, New Haven.

Acknowledgment: We wish to thank Fred Staats for assistance in testing subjects.

Essentially this procedure involves presenting simultaneously an overload of items, usually letters or digits, in a very brief tachistoscopic exposure which is followed after a similarly brief period of time by a probe or indicator designating which element or subset of elements the subject has to report. Despite the fact that the display load generally exceeds the memory span, if the indicator occurs soon enough after the display the subject can give a highly accurate report of the specific element(s). As demonstrated by Sperling (1960) this delayed partial-sampling procedure shows the subject has far more information available than can be reported by the memory-span, or whole-report, technique. Presumably the information tapped by the partial report exists in a storage medium of such brevity that the memory-span, or whole-report, technique is too slow to reveal it. The superiority of partial report over whole report declines rapidly with delay of indicator. Estimates of the decay time of iconic memory inferred from the decline in accuracy of partial reports are of the order of 250 msec to several seconds depending on the prevailing luminance conditions (Averbach and Coriell, 1961; Averbach and Sperling, 1961; Keele and Chase, 1967).

The proposition that iconic memory is literal, or precategorical, receives support from the sorts of selection criteria which allow for efficient performance in the delayed partial-sampling task. In the original experiments of Sperling (1960) subjects were presented with an array of several rows of letters or digits. The delayed indicator specified report by row or column. Partial report at brief delays of the indicator was superior to whole report demonstrating that the spatial properties of the input were available in the iconic representation. However, in one experiment Sperling used a stimulus array consisting of letters and digits intermixed and cued for partial report by category, i.e., report the letters or report the digits. In this instance partial report was not superior to whole report, suggesting the distinction between letters and digits is not available at the level of iconic memory. Such a distinction is based on a derived property of the stimulus. Presumably the time required to categorize a particular set of physical characteristics as representing an item belonging to the class "letters" or "digits" is considerable in the context of iconic memory. In contrast, superior partial report over whole report can be clearly demonstrated when the criterion for selection is brightness, size (Von Wright, 1968), color (Clark, 1969; Von Wright, 1968), shape (Turvey and Kravetz, 1970), or as already indicated, location (e.g., Sperling, 1960). These data demonstrate that we are able to select or ignore items in iconic memory on the basis of their general physical features. We cannot, however, with the same efficiency select or ignore items on the basis of their derived properties. All this speaks to the precategorical nature of iconic memory.

The investigation of the analogous sensory register in the auditory system has been conducted along somewhat different lines. The starting point for one approach (Crowder and Morton, 1969) has been the pronounced recency effect in the recall of serial lists presented in the auditory mode. That the recency effect is tied to the auditory modality and not to a subsequent communal short-term categorical store is shown by the fact that it is abolished by a redundant auditory suffix (Crowder and Morton, 1969) but not by a redundant visual suffix (Morton and Holloway, 1970). Moreover, this suffix effect does not occur with delay of the suffix beyond 2 seconds (Crowder, 1969). These results serve to define a brief auditory store which lasts a little longer than iconic memory. Moreover, it appears that material is held in this store in a form in which only relatively crude attributes of

the stimulus are distinguished. Whereas the conceptual class to which the suffix belongs is unrelated to the size of the suffix effect (on recency), the physical channel over which the suffix occurs (voice quality, spatial location) is important. Nonspeech suffixes have no effect.

These results suggest that the information is not stored in an alphanumerically categorized form, although it could perhaps be handled, if less elegantly, without recourse to such a prelinguistic store.

An auditory analogue to the Sperling procedure could provide a "converging operation" on the problem of the form of auditory storage.

Moray, Bates, and Barnett (1965) have shown that after multichannel auditory stimulation, partial report of one channel is relatively superior to whole report. Although their experiment used from one to four items on up to four different input channels, only one time delay was used and only one mode of recall (spatial location). It is thus not clear whether the superiority they obtained for partial report is simply attributable to output interference.

The experiments reported here use a similar paradigm to that employed by Moray et al. but explore the effect of time delays on both partial and whole report and of requiring report by spatial location and/or by category.

EXPERIMENT I

This experiment presented two consecutive digits simultaneously on each of three different channels and cued partial report with a brief tone 0, 1, 2, or 4 seconds after the stimuli.

Method

The basic material for this experiment was three sets of the digits, one through ten, synthesized on the Haskins parallel formant synthesizer after retouching the output from a speech-synthesis-by-rule program (Mattingly, 1968). Each set was synthesized at a different fundamental frequency; two sets had formant values appropriate to a man's voice, and the third, values appropriate to a woman's. The fundamentals were within the appropriate adult sex ranges, and each digit lasted 250 msec. Using a computer-controlled pulse-code modulation system (Cooper and Mattingly, 1969) which mixed appropriate digits and simultaneously output two different signals, an experimental tape was made up so that on each trial the subject heard six different digits over stereophonic headphones. These digits appeared to come from three different spatial locations, left, middle, and right, the two digits on any one channel arriving simultaneously with the two digits over the other two channels. Throughout the whole experiment the same voice appeared on the same channel (the female voice was in the middle).

Twenty different six-digit combinations were used, each combination appearing in the same spatio-temporal pattern. At some time after the six digits a 19-msec, 2-kHz tone appeared at one of the three locations. This tone came either immediately after the end of the stimulus or after a 1-, 2-, or 4-second delay. Each digit combination appeared once with the twelve possible tone conditions (3 locations x 4 delays) to give a basic set of

240 trials. These were randomized into ten blocks of 24 trials each so that each half block had one of each tone condition and each block at least one and no more than two of the same digit combination. The intertrial interval between a tone and the next stimulus was 10 seconds.

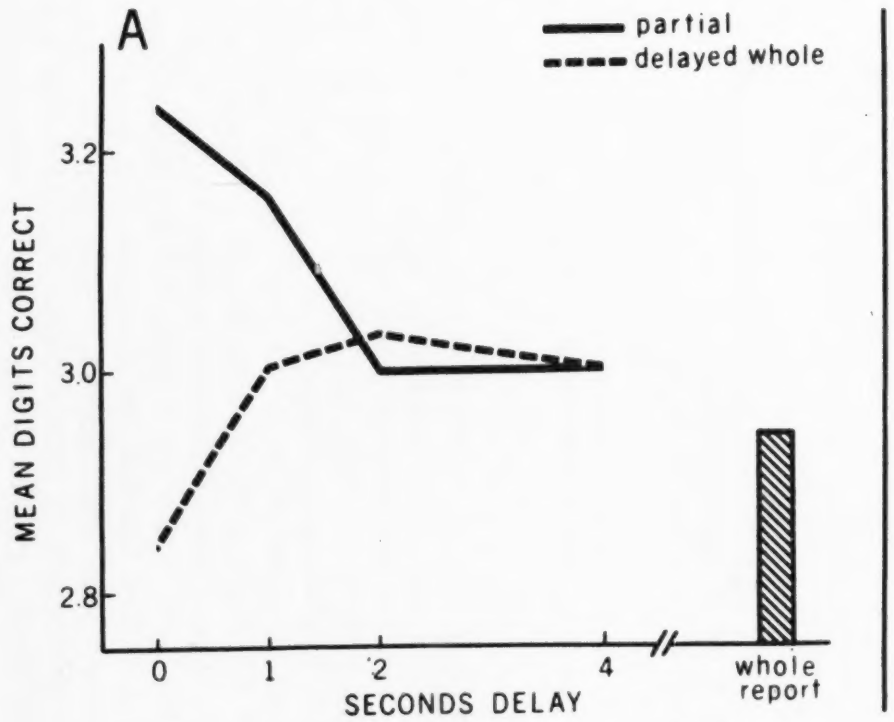
The subjects (four Yale undergraduates) heard this tape under three different instruction conditions: whole report, partial report, and delayed whole report. For whole report they had to write down all six numbers as soon as they wished, being sure to write each number in a column which assigned it to its correct spatial location. They were told (truthfully) that a correct digit in the wrong location would be scored as wrong. For the partial-report condition they had to wait for the tone and then write down the two digits which occurred on that channel. For the delayed whole report they again had to wait for the tone and then write down all the digits, starting with those on the tone's channel, attributing them to their correct locations. For all three conditions they were told that within a particular channel it did not matter in what order they wrote down the two digits. In this and subsequent experiments they were told to guess, if not sure, to make up the required number of partial- and whole-report items.

All subjects started with five blocks (each of twenty-four trials) of whole report. They then took a total of twenty blocks of partial report and ten blocks of delayed whole report, alternating ten on partial with five on delayed whole. Three subjects started this alternation with partial and one with delayed whole. All subjects finished with a further five blocks of whole report. They were tested individually in a soundproof chamber in about four sessions of about an hour each spread over a two-week period.

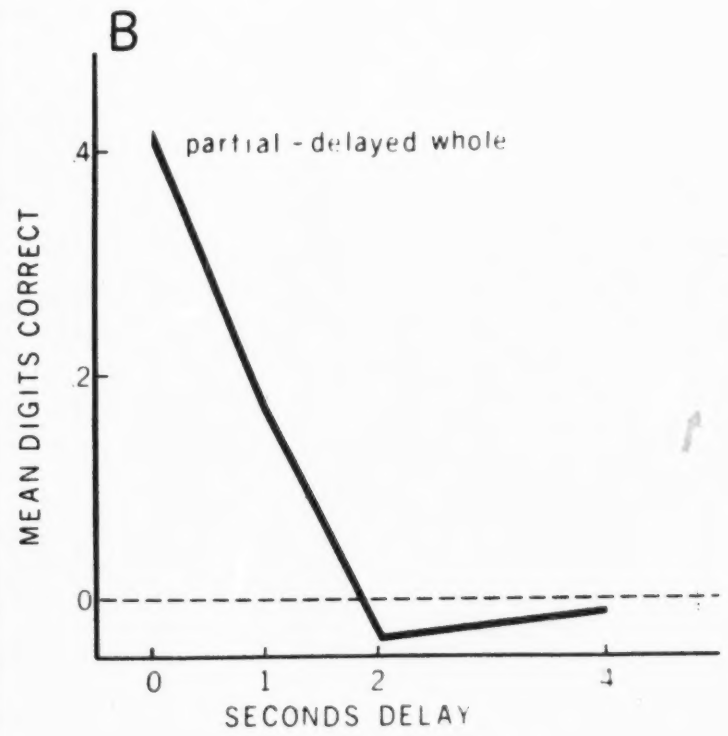
Results and Discussion

Responses were scored as correct in the two whole-report conditions only if they were attributed to their correct channel. For the partial report the scores were multiplied by three for comparison with the whole-report scores. The results are displayed in Figure 1a. The bar on the right of the figure shows the whole-report performance (2.94 items). The partial- and delayed whole-report conditions are shown as a function of the delay of the indicator. The partial-report scores decayed to and the delayed whole-report scores rose to an asymptotic value of about three items, both reaching this value by about 2 seconds. Figure 1b shows the difference between the partial- and the delayed whole-report scores, again as a function of indicator delay. When the indicator was given immediately after the stimulus, partial report was superior to "delayed" whole report by about 0.4 items. With a one-second delay this difference was less than 0.2 items, and by 2 seconds it had vanished. An analysis of variance on the partial- and delayed whole-report scores showed a significant interaction of indicator delay with report condition [$F(3,9) = 6.8$; $p < .025$)]. The two main terms were not significant. Separate analyses of variance on the partial- and delayed whole-report scores showed only marginal evidence ($p < .1$) for their change with time. Some justification is therefore needed for taking the delayed whole-report condition, rather than the regular whole-report, as the relevant comparison for partial report. To minimize any masking of the stimuli, the indicator was presented at a relatively low intensity. This made it rather difficult to detect, particularly in the zero-delay condition. Compared with the whole-report condition in which the subject could ignore the tone, the partial-report condition imposes an additional perceptual burden,

Partial and Delayed Whole Report by Spatial Location as a Function of Auditory Indicator Delay in Experiment I



Difference Between Partial and Delayed Whole Report as a Function of Indicator Delay



Note: Bar on right is whole report: indicator present but ignored. Maximum possible score is six items.

Fig. 1

which possibly impairs perception of the digits. In the delayed whole-report condition, however, the subject has to perform exactly the same operation on the indicator (detect it and identify which channel it is on) as in partial report. Particularly for short indicator delays then, the partial-report condition may be underestimating the subjects' abilities relative to the whole report, but not relative to the delayed whole report.

The increase in delayed whole report with increasing indicator delay resembles the finding by Crawford, Hunt, and Peak (1966) that recall of material that could be "chunked" improves as more time is allowed for organization before recall. Similar principles may be operating here, although perhaps at a different level of organization (cf., Hunt, 1971).

Although the results of this experiment are complicated by the variation in delayed whole-report scores and the magnitude of the advantage for partial over whole report is small (.4 item), there is sufficient indication of some form of transient memory to warrant further work.

EXPERIMENT II

The next experiment incorporates the following changes: (a) a visual indicator to avoid some of the difficulties encountered with the auditory one; (b) natural speech, intentionally poorly synchronized across channels, to help channel separation; (c) nine items instead of six to attempt to increase the magnitude of the effect; and (d) a mixture of digits and letters.

Method

The nine numbers, one through ten (omitting disyllabic 7), and the nine letters, BFJLMQRUY, were randomly assigned to twenty, nine-item stimuli (three items on each of three channels) with the following restrictions: (1) each channel of each stimulus contained two items of one category and one item of the other; (2) each stimulus had four items of one category and five of the other; (3) each category was equally represented over all twenty stimuli; (4) between stimuli, each position on each channel contained each item at least once.

This then gave a set of twenty main stimuli, each of which had three "mini-lists" of three items corresponding to a particular channel. For example, one of the stimuli had BU3 on the left channel, 52J on the middle, and 8F6 on the right. Each of these sixty, three-item mini-lists was recorded as a single continuous utterance by a native speaker of British English at a rate of three items per second. These recordings were then assembled into a tape similar to that used in the first experiment, but without any auditory indicator. One other difference was that a tone appeared 3/4 second in front of the stimulus to act as a warning to the subjects and to trigger the timing apparatus. The interstimulus interval on the tape allowed for a 10-second pause between the indicator and the next stimulus. The assignment of partial report conditions to the twenty stimuli was done in the same way as in the last experiment, giving ten blocks of twenty-four trials. Only partial and regular whole report was used.

The indicator was a slide with a vertical black bar on the left, middle, or right, which was projected onto a screen in front of the subjects. The

subjects were twelve Yale undergraduates who were tested in groups of four in a sound-treated room for one session of about 3 hours. Their instructions were similar mutatis mutandis to those of the last experiment. They were urged to keep an eye on the screen while they heard the stimuli so that they could react to the indicator as soon as it occurred. Again, as in the previous experiment, partial- and whole-report conditions were given in separate blocks. In the whole-report condition no visual indicator was used, but subjects knew in advance which condition to expect. All the subjects had initial practice of one block of whole report followed by four blocks of partial report. They then took a total of five blocks of whole and ten of partial, alternating as in the first experiment, with the condition that started the alternation balanced across subjects. Only these last fifteen blocks were scored.

Results and Discussion

Two variables are interesting in the data: the effect of indicator delay and the position of an item within a particular "mini-list." Figure 2 shows the data as a function of these two variables. Observe first that for each item position the partial-report curves descend towards the whole-report level but that the absolute level of each curve and its corresponding whole report varies. In the analysis of variance this was reflected in highly significant main terms for indicator delay [$F(3,99) = 3.91$; $p < .001$] and item position [$F(2,22) = 33.22$; $p < .001$]. Individual t-tests showed that the third item was recalled significantly better than the other two ($p < .001$) which did not differ between themselves ($p > .1$). The interaction between the two main variables was not significant ($F < 1.0$)--there was no change in decay with item position, so the three curves have been condensed into a single curve in the inset of Figure 2. A separate analysis of variance with the 4-second delay condition vs. whole report as one factor and item position as the other factor gave no significant main effect or interaction, so we are justified in including the average whole-report bar on this figure.

Separate t-tests on the difference between the average values of the four delay conditions and the average whole-report value gave highly significant differences for 0, 1, and 2 seconds' delay ($p < .001$), all twelve subjects showing the effect for all three conditions. As suggested above there was no significant difference between the 4-second delay condition and whole report [$t(11) = 1.6$; $.1 > p > .05$], eight subjects showing superior partial report and four showing superior whole report. Whatever, then, is responsible for superior partial over whole report loses its effect if the indicator is given 4 seconds after the end of the stimulus.

The other interesting feature of the data is the effect of item position in each "mini-list." The last item was better recalled than the first two. This cannot be attributed to response bias since the stimuli were so constructed that each item was about equally represented at each list position. Nor can this advantage for the last item be attributed simply to a shorter time elapsing between the item being presented and the indicator appearing. To see this, look again at Figure 2. Compare for example the performance on the third item at 1-second delay (5.34 items) with that on the second item with zero delay (4.36 items). Here the time elapsing between stimulus and indicator is greater for the condition that shows better recall. The relevant difference appears to be that in the case of the second item there is another

Partial Report by Spatial Location as a Function of Visual Indicator Delay in Experiment II

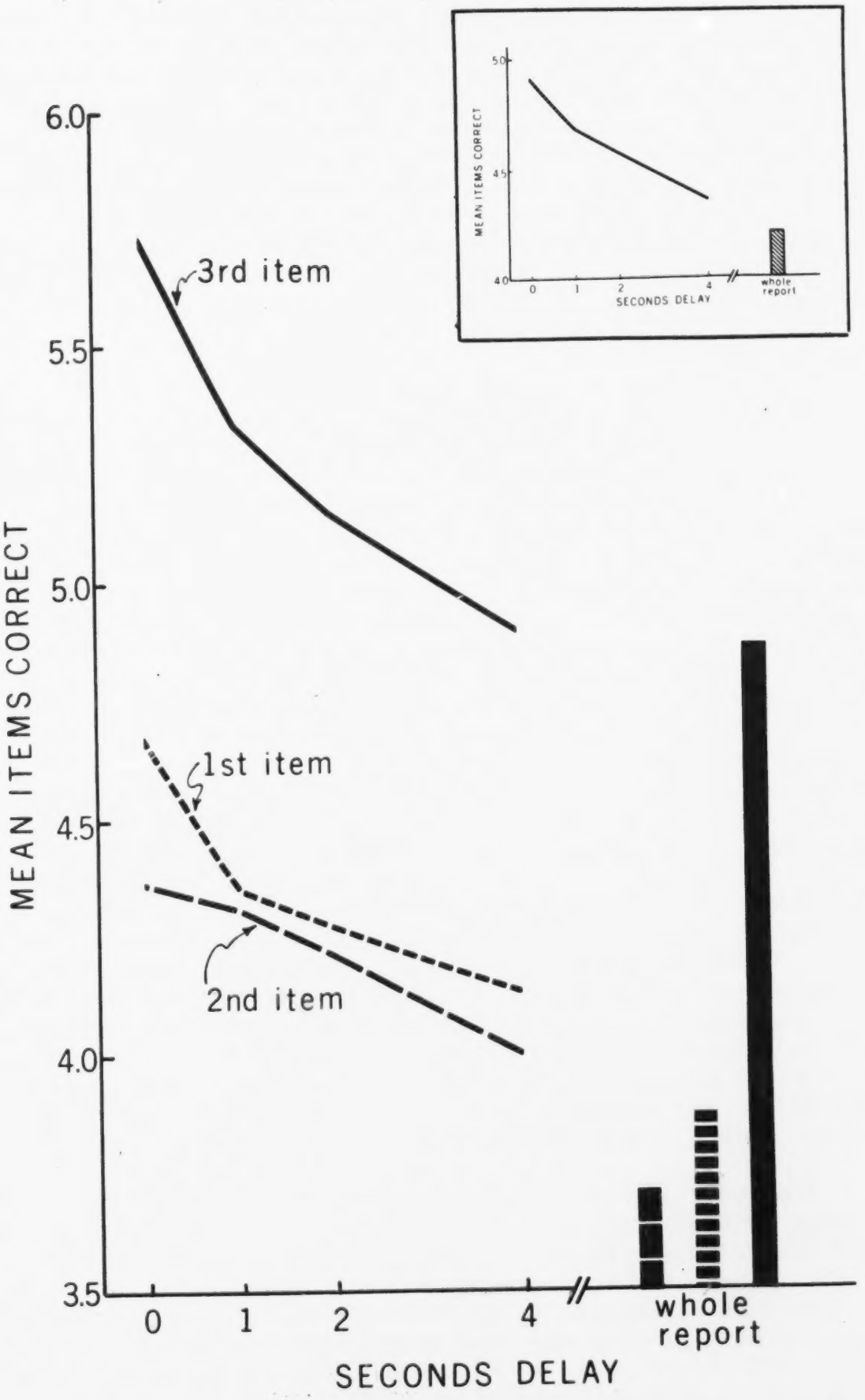


Fig. 2

Note: The three curves correspond to the temporal order of the three items on a channel. The bars on the right are the whole reports, made with no indicator. The inset is the average of the three curves with its average whole report. Maximum possible score is nine items.

item presented immediately after it, while for the third item there is a helpful silence (Aaronson, 1968). By contrast, the second item does not exert a similar effect on the first item, which is in fact recalled insignificantly better than the second. It is unlikely that these effects are due to the intelligibility of the individual sounds, since the editing procedure used to make each mini-list the same duration tended to chop off the extremes of the list and so to impair the first and last items relative to the middle one. Others have reported marked recency effects in the recall of the unattended channel in dichotic listening experiments (Bryden, 1971; Murray and Hitchcock, 1969). Murray and Hitchcock find, using a probe technique, that recall of the last item on the unattended ear is markedly superior to that of previous items. We would, however, doubt the validity of inferring a specific decay time for auditory memory from this result since, as we have shown here, the presence of an interpolated item is more detrimental to recall than is a delay of more than 2 seconds. Both interference and decay appear to be potent factors in the temporal degradation of auditorily input material.

Why was the magnitude of the advantage for partial report so small compared with the large advantages evident in the visual case? The most plausible reason is uncertainty as to where items had occurred. Many subjects expressed difficulty in hearing the middle channel as a separate source and, although no tests were run with the cue given well in advance of the stimuli, it is very likely that a considerable number of errors would have been made. By contrast, in the visual modality performance is very good with precuing by location (Eriksen and Collins, 1969). If the stimuli are difficult to distinguish along the dimensions used to cue the partial report, selection would be less efficient than if they were readily distinguished. Greater superiority for partial over whole report may be obtainable with different voices on the three channels to make them more discriminable.

EXPERIMENT III

The results of Experiments I and II functionally define a store in which material is held for about 2 seconds, though the form in which the material is held remains unclear. From the data presented here there is, in fact, no direct evidence that the store is specific to auditorily presented material. Averbach and Sperling (1961) report, for example, that partial report of visual material remains superior to whole report for longer than 2 seconds if dark pre- and postfields are used. With light fields they find a much more rapid decay on the order of 250 msec, so the store that they identify must have some component which is sensitive to the purely visual parameters of the stimulus situation. Unfortunately, we have no evidence that the store we have identified for auditorily presented material is similarly restricted by auditory stimulus parameters.

A lever that has been applied to this question for the visual store can also be applied here. Partial report is superior to whole report in the visual case only when report is cued along some "physical" dimension of the stimulus. Recall by higher-order categories shows very little advantage for partial over whole report, and this, as noted above, has been taken to imply that the items are not classified by higher-order categories in the iconic store. We can ask a similar question in the auditory case. Does recall by category give an advantage for partial over whole report similar to that obtained for recall by spatial location? The next two experiments provide some data on this question.

Method

Part of the tape for Experiment II was used with eleven new subjects. This time, however, only two different indicators were used. A vertical bar to the left indicated that only the numbers were to be recalled and a bar to the right, only the letters. Because of this reduction in partial-report conditions, only two-thirds of the original tape was used, with whole and partial report distributed in the same proportion as in the last experiment. The subjects were given five blocks of practice as before.

In this partial-report condition they were told to write down only the particular category denoted by the indicator and to put the items down in their correct location. In effect they were answering the question "What were the numbers, and where were they?" In the whole-report condition they were given identical instructions to those of the previous experiment. They did not have to recall the items in any particular order, so long as they attributed each item to its correct location.

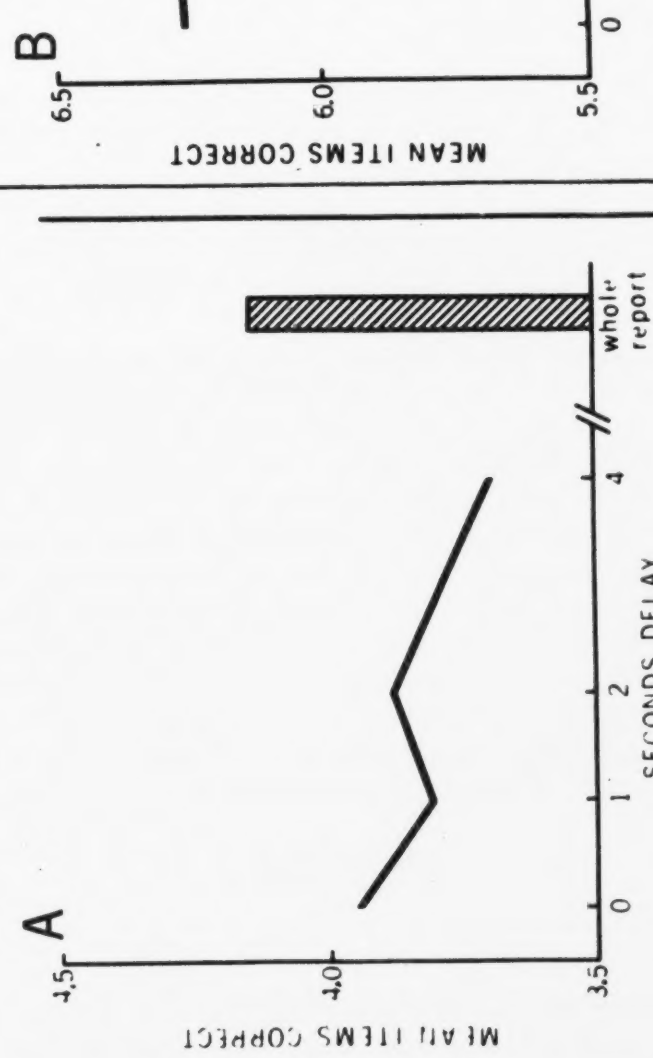
Results

The analysis of variance showed a significant main effect of item position in a list as in the previous experiment [$F(2,20) = 19.16$; $p < .001$]--again with #3 superior to #1 and #2--but no main effect of indicator delay [$F(3,30) = 2.18$; $p > .1$] nor any significant interaction between it and item position [$F(6,60) = 2.20$; $.1 > p > .05$]. Figure 3a shows the data averaged over item position. The striking difference between this figure and Figure 2 is that, although the whole report was almost identical to that of Experiment II (4.2 items), partial report tended to be lower than whole report. This is significant for the average of the partial report conditions [$t(10) = 2.36$; $p < .025$]. Clearly report by category is, in this situation, an inefficient mode of recall, and one that shows little variation with indicator delay.

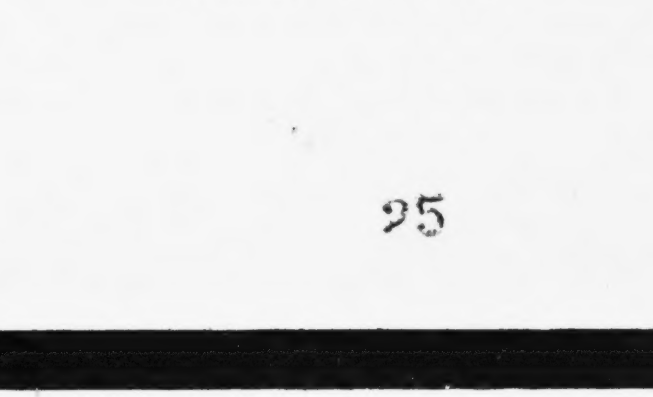
Discussion

These results raise two questions: Why was there no change in partial report with time? and Why was partial report less efficient than whole report? The absence of decay with time, of course, supports the hypothesis that semantic category information is not available in the store whose decay leads to the decline in partial report. But, unfortunately, other reasons could be advanced for the absence of temporal decay. It could be that, since the transformation between the indicator and the particular selection required is more complex in this experiment than in the previous one, the subject requires more time to perform it and consequently can tap the transient memory only after it has suffered considerable decay (cf., Eriksen and Collins, 1969). Against this we can offer informal observations on the subjects: they generally started to make their response at least within a second of the indicator appearing on the screen and, when questioned after the experiment on the difficulty of identifying the required category, considered it a trivial imposition. Doubtless it required time and effort in the early stages of the experiment, but it probably became automatic by the end of the practice period of 120 trials. Almost no errors were made in deciphering the indicator, and they were scored as if the chosen category were the correct one. Comparison of visual partial-report procedures (Averbach and Coriell, 1961; Sperling, 1960) shows that for trained subjects the estimate of iconic storage decay

Partial and Whole Report by Letter/Digit Category
when Location is not Required in Experiment IV



Partial and Whole Report by Letter/Digit Category
when Location is Required in Experiment III



Note: Maximum possible score is nine items.

Fig. 3

does not depend on whether the indicator used is visual or auditory, although undoubtedly the auditory indicator involves a more complex transformation than the visual and the estimated decay time is very much shorter than here. A logical objection to the problem of when the indicator information is used must remain, but we doubt that it is of great practical significance.

A more substantial objection is that partial report by category required recall of more material (4.5 items) than recall by channel (3 items). To see the effect of this increase in partial-report size, consider the extreme case of "partial" report of all nine items. Here, of course, there would be no advantage for partial report over whole since they are identical. Neither, by analogy with the visual case (Averbach and Sperling, 1961), would there be any decay with time. The absence of any significant decay in partial report in this experiment could be attributed to this factor. As a counter to this argument, we can compare the partial report curves obtained by Sperling (1960) and Averbach and Coriell (1961); the former required partial report of three or four letters respectively from a nine- or eight-letter array, whereas the latter required report of only one letter from a sixteen-item array. Despite this large difference in the fraction of material required in the partial report, similar estimates of the useful life of iconic memory were derived. The comparison with vision here may not be valid since, for example, read-out times from the transient store into a more permanent form may be more rapid in vision than in audition.

Why was partial report worse than whole report? In both conditions subjects had to assign items to their correct spatial location, all nine items in the whole report but only those of a particular category in the partial. For the partial condition, an explicit decision about an item's category is required, a decision which is not necessary for the whole report. Paradoxically, subjects would have done better to have written a whole report and then deleted the inappropriate category, since conscious omission appears to involve effort which can impair memory of items not already committed to paper. For partial report of more than one item this extra cognitive load could influence the decay of partial report since items would be output at a slower rate (cf., Posner and Mitchell, 1967). Perhaps, then, this extra interrogation of an item's category is responsible for the absence of a decline in partial report with time, not because category information is not available in the store responsible for the decline but because read-out from that store is slower when information on two attributes of an item (location and category) is required rather than information on just one (location). If this were in fact the case, and category information were as accessible as location information, we would expect to find decay of partial report, when cued by category with the location of an item, irrelevant. Accordingly, the next experiment looks at recall by category when location is not required.

EXPERIMENT IV

Method

The same tape and slides were used as in the last experiment. Eight new subjects were given instructions and practice similar to those of the previous experiment the only difference being that they were not told to remember or report the location of a particular item. Their answer sheets were divided into two columns; for the whole report they wrote the numbers on the left and

the letters on the right, for the partial they wrote the cued category in the left-hand column. They were told to average about 4 1/2 responses per trial for the partial report and to write all nine items for the whole report, guessing if necessary.

Results

Partial report is now at approximately the same level as whole report. No breakdown of the scores was made in terms of item position, and the average of the three positions is shown as a function of indicator delay in Figure 3b. The analysis of variance showed a significant effect of indicator delay [$F(3,21) = 3.97$; $p < .025$]. However, the magnitude of the effect was very small; the difference between the whole report and the zero-delay partial condition was only 0.25 items. This was significantly smaller than the 0.71 items found in Experiment II for partial recall of the same material by location rather than by category [$t(18) = 5.73$; $p < .001$]. Partial report was significantly greater than whole report for the zero and 1-second delay conditions ($p < .01$) but not for the 2- and 4-second conditions ($p > .1$).

Discussion

When partial report was required by semantic category, there was some advantage over whole report for indicator delays of a second or less. However, this advantage was significantly less than when partial report was required by spatial location. As we suggested in the discussion of the previous experiment, the magnitude of the partial-report advantage over whole report depends *aliis aequis* on the relative number of items required for partial report. We cannot tell on the evidence presented here whether the much smaller advantage under recall by category is due to the larger number of partial response items or to the relative ease of withdrawing items from a decaying store according to different stimulus attributes.

The lower partial report over whole report obtained in Experiment III is clearly not attributable simply to the fact that recall was cued by category. Rather it must be due to the fact that recall required memory for two attributes of the stimulus rather than one. The small, though significant, decay found in this experiment suggests that this may also have been responsible for the absence of any decay in Experiment III.

GENERAL DISCUSSION

The evidence presented in these four experiments demonstrates some transient memory for auditorily presented material, from which we have reason to believe retrieval is more easily made according to the dimension of physical location than according to an item's semantic category. The time limit on the store identified in these experiments is similar to that reported from other experiments in audition. Treisman (1964) reports that the identity of two messages dichotically presented is noticed when the non-shadowed message leads only if the temporal disparity is less than about 1 1/2 seconds. When the shadowed message leads, the critical time is around 4 seconds. One disadvantage of the design of our experiments is that there is no attempt to control, as in Treisman's experiments, the attentional strategy of the subject. Nevertheless our figure of something greater than 2 seconds but less than 4 is conveniently bracketed by Treisman's

two estimates. It is perhaps more likely that our store has more in common with the 1 1/2-second condition in Treisman's experiments since there was a silent interval after our to-be-remembered sounds which perhaps extended its useful life. Glucksberg and Cowen (1970) give a figure of less than 5 seconds for memory for digits embedded in prose on the rejected channel of a shadowing task, a figure compatible with a similar experiment by Norman (1969), which used a string of six digits rather than an embedded digit. They comment also that their subjects were never aware that a digit had occurred unless they could name the particular digit; they had no general awareness of the occurrence of a member of the class of items required, neither were context effects any help in detection. These observations correspond well with Treisman's findings and with our own findings of less efficient partial report by semantic category than by spatial location.

The presumed absence of semantic attributes, however, cannot serve to distinguish between material held in some articulatory/phonetic code and material held in some less processed auditory form. The only argument in favor of the latter, and it is not a strong one, is that the lower limit on the detection of periodicity for repeating white noise is of the order of 1 second (Guttman and Julesz, 1963), a time which is not incompatible with our estimate, considering the finer auditory resolution required to distinguish two sections of statistically identical white noise compared with that required to distinguish between eighteen acoustically very different items. It is perhaps significant that in the first experiment reported here, which used synthetic speech and probably thus required better auditory resolution, the critical time was apparently less than 2 seconds.

REFERENCES

Aaronson, D. (1968) Temporal course of perception in an immediate recall task. J. Exp. Psychol. 76, 129-140.

Averbach, E., and Coriell, A.S. (1961) Short-term memory in vision. Bell System Technical Journal 40, 309-328.

Averbach, E., and Sperling, G. (1961) Short-term storage of information in vision. In C. Cherry (ed.), Information Theory: Proceedings of the Fourth London Symposium. (London, Butterworth).

Bryden, M.P. (1971) Attentional strategies and short-term memory in dichotic listening. Cog. Psychol. 2, 99-116.

Clark, S.E. (1969) Retrieval of color information from the pre-perceptual storage system. J. Exp. Psychol. 82, 263-266.

Cooper, F.S., and Mattingly, I.G. (1969) Computer controlled PCM system for investigation of dichotic speech perception. J. Acoust. Soc. Amer. 46, 115(A).

Crawford, J., Hunt, E., and Peak, G. (1966) Inverse forgetting in short-term memory. J. Exp. Psychol. 72, 415-422.

Crowder, R.G. (1969) Improved recall for digits with delayed recall cues. J. Exp. Psychol. 82, 258-262.

Crowder, R.G., and Morton, J. (1969) Pre-categorical acoustic storage (PAS). Percept. Psychophys. 5, 365-373.

Eriksen, C.W., and Collins, J.F. (1969) Temporal course of selective attention. J. Exp. Psychol. 80, 254-261.

Glucksberg, S., and Cowen, G.N., Jr. (1970). Memory for non-attended auditory material. Cog. Psychol. 1, 149-156.

Guttman, N., and Julesz, B. (1963) Lower limits of periodicity analysis. J. Acoust. Soc. Amer. 35, 610.

Haber, R.N. (1969) Information processing analysis of visual perception: An introduction. In R.N. Haber (ed.), Information Processing Approaches to Visual Perception. (New York, Holt, Reinhart and Winston).
Hunt, E. (1971) What sort of a computer is man? Cog. Psychol. 2, 57-98.
Keele, S.W., and Chase, W.G. (1967) Short-term visual storage. Percept. Psychophys. 2, 383-386.
Mattingly, I.G. (1968) Experimental methods for speech synthesis by rule. IEEE Trans. AU-16, 198-202.
Moray, N., Bates, A., and Barnett, I. (1965) Experiments on the four-eared man. J. Acoust. Soc. Amer. 38, 196-201.
Morton, J., and Holloway, C.M. (1970) Absence of a cross-modal "suffix effect" in short-term memory. Quart. J. Exp. Psychol. 22, 167-176.
Murray, D.J., and Hitchcock, C.H. (1969) Attention and storage in dichotic listening. J. Exp. Psychol. 81, 164-169.
Neisser, U. (1967) Cognitive Psychology. (New York, Appleton-Century-Crofts).
Norman, D. (1969) Memory while shadowing. Quart. J. Exp. Psychol. 21, 85-93.
Posner, M.I., and Mitchell, R.F. (1967) Chronometric analysis of classification. Psychol. Rev. 74, 392-409.
Sperling, G. (1960) The information available in brief visual presentations. Psychol. Mono. 74, no. 11.
Treisman, A.M. (1964) Monitoring and storage of irrelevant messages in selective attention. J. Verb. Learning and Verb. Behav. 3, 449-459.
Turvey, M.T., and Kravetz, S. (1970) Retrieval from iconic memory with shape as the selection criterion. Percept. Psychophys. 8, 171-172.
Von Wright, J.M. (1968) Selection in visual immediate memory. Quart. J. Exp. Psychol. 20, 62-68.

Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening

C.J. Darwin*
Haskins Laboratories, New Haven

The origin of the ear asymmetry effect in dichotic listening has been attributed to a number of factors: perception (Kimura 1961a,b), memory (Inglis, 1962; Oxbury, Oxbury, and Gardiner, 1967) and attention (Treisman and Geffen, 1968) all being espoused as candidates. Recent work on the recall of dichotically presented nonsense syllables has reinforced Kimura's original hypothesis that the effect originates in a difference in the efficiency with which the two hemispheres perceive auditory material (Darwin, 1971a; Haggard, 1971; Studdert-Kennedy and Shankweiler, 1970). Whether or not a right-ear advantage appears is not entirely predictable simply in terms either of the acoustic features in the stimulus presented to the subject or of the phonetic response category to whose perception they contribute, but rather in the relationship between the two (Darwin, 1971a). If short-term memory variables were paramount in determining the ear difference effect, we would expect the phonetic category to be the only relevant variable. This paper looks at the question of stimulus and response factors as determinants of the right-ear advantage from a slightly different angle, that of determining the conditions of contralateral stimulation under which the effect occurs.

It is generally true that ear differences are obtained more readily under conditions of dichotic presentation than monotic. The most convincing evidence of this comes from work on commissurotomized patients who can report perfectly digits presented to the left ear when only the left ear is stimulated but can report very little from the left ear when different digits are played simultaneously to the opposite ear (Milner, Taylor and Sperry, 1968). The extent of the suppression of the left ear depends on the clarity of the signal on the right ear; the greater the distortion, the less from the left ear is recalled (Sparks and Geschwind, 1968).

Work with normal subjects also shows that the nature of the competing stimulus is important. In otherwise similar paradigms, Kirstein and Shankweiler (1969) find a reliable right-ear advantage for stop-vowel syllables when they are opposed by another stop-vowel syllable, whereas Corsi (1967) failed to find any right-ear advantage for nonsense syllables opposed by white noise. Darwin (1971b) also found no right-ear advantage for stop-vowel syllables opposed by a random noise. Contralateral white noise, however, can enhance the ear difference for a two-click threshold task (Murphy and Venables, 1970). The relationship between the signals on the two ears, rather than the absolute nature of the signal, thus appears to be the important variable.

For ear differences to be revealed there must be both a relevant difference between the hemispheres and sufficient functional decussation to ensure

---

*Now at the University of Sussex.

that each ear is projected predominantly to the opposite hemisphere. Without a competing stimulus, that is, under monaural stimulation, the latter condition is presumably not satisfied (Kimura, 1961a,b). If such competition is necessary to ensure adequate functional decussation, we might presume that this competition is only necessary in principle up to the stage of processing at which the hemispheres become functionally distinct, since after this stage the two input signals can be distinguished by this treatment rather than simply by their ear of arrival. If decussation were not sufficient up to this first stage of hemispheric differentiation, no distinction between the hemispheres at this or any subsequent stage would be detectable in the response.

If, indeed, differences between the hemispheres first appear at some perceptual level, rather than at a level associated with the organization of the response, we might expect that, for ear differences to be obtained, the competing stimulus need only be in the same perceptual class as the stimulus on the other ear, rather than actually in the response class used in the experiment.

Method

The response set for this experiment consisted of the stop consonant syllables /ba, pa, ga, ka/. One of these sounds was always present on the ear which the subject attended and was asked to report. The other ear received one from a set of sounds. Which set was used constituted the experimental condition. The three sets were

1. /ba, pa, ga, ka/ (same as response set)
2. /ba, pa/ (two sounds from response set)
3. /da, ta/ (two sounds not in response set but in same perceptual class).

In the first set each sound on the attended ear was paired with every other sound in the response set except itself an equal number of times. The second and third sets, however, were restricted so that the two sounds in a dichotic pair always differed in voicing. The sounds used were prepared on the Haskins parallel formant synthesizer and assembled into a dichotic tape using a special computer program (Mattingly, 1968).

The experimental tape contained one block of forty-eight trials for each stimulus condition. One ear was attended for half a block and then the other ear for the remainder of the block. Each block was taken twice in each of two headphone orientations by each subject. The ordering of the blocks and which ear was attended was approximately counterbalanced over subjects (four of the six block orderings had six subjects and two had four). Thirty right-handed undergraduates, none of whom had previously taken part in a dichotic listening experiment, participated in the experiment.

The subjects were introduced to the sounds of the response set and given some practice at identifying them singly. Only those who did better than 75 percent correct on the single sound identification were allowed to proceed to the dichotic test. For the dichotic test subjects were told that they would get two different sounds, one in each ear, that they were to attend to

a given ear for a sequence of twenty-four trials, and that the sound in that ear would always be one of /ba, pa, ga, ka/, although the sound in the other ear might be something else. They were not told which stimulus condition they would receive nor that the sounds /da, ta/ were also being used in the experiment.

Results

Because of the restriction on the voicing dimension for stimulus conditions 2 and 3, the results were scored only for place of articulation; voicing of both the stimulus and the response was made irrelevant. Confusion matrices (2x2) were constructed for each subject and each stimulus condition, and from these matrices simple percent correct scores were calculated along with a measure of the discriminability of place of articulation untainted by variations in response bias between stimulus conditions. The measure used is derived from Luce (1959) and is

$$\log_e \alpha = \frac{1}{2}[p(R1/s1)p(R2/S2)/p(R2/S1)p(R1/S2)].$$

This measure is almost identical to the d' of signal detection theory but is more readily applicable to larger matrices and is computationally more convenient (Haggard, 1968).

The percent correct values appear in Table 1 and the $\log_e \alpha$ in Table 2.

Table 1. Percent correct for recall of place of articulation of the stops /b,p,g,k/ when opposed by different stimulus sets.

Ear	Opposing Set			
	bpgk	bp	dt	Mean
left	64.6	86.0	69.7	73.4
right	71.9	90.7	74.0	78.9
left + right	68.3	88.3	71.9	76.1
right - left	7.4	4.7	4.2	5.4

Table 2. Mean $\log_e \alpha$ for recall of place of articulation of the stops /b,p,g,k/ when opposed by different stimulus sets.

Ear	Opposing Set			
	bpgk	bp	dt	Mean
left	.366	1.238	.670	.758
right	.586	1.420	.822	.942
right + left	.476	1.329	.746	.850
right - left	.220	.182	.152	.184

Friedman two-way analyses of variance across all three stimulus conditions were not significant for right minus left ear scores either for the percent correct ($p > .2$) or for $\log_e \propto$ ($p > .1$). There was thus no significant variation in ear advantage over the three stimulus conditions. Combining the data from all three conditions gave a significant right-ear advantage for the percent correct score ($p < .01$; 1-tailed Wilcoxon) and for the $\log_e \propto$ ($p < .005$; 1-tailed Wilcoxon). The performance level (right plus left ear scores), however, showed highly significant variation on a Friedman test between the three stimulus conditions on $\log_e \propto$ ($p < .001$). There was a significant difference in right plus left ear scores between stimulus condition 2 and both 1 and 3 ($p < .001$). The subjects are thus reflecting some aspect of the change in stimulus condition.

Since the first stimulus condition is somewhat peripheral to the main question asked by this experiment, a Wilcoxon T-test was used to test whether the ear difference is any larger for condition 2 than for condition 3. This showed a quite insignificant trend in the opposite direction. Combining conditions 2 and 3 gave a significant right-ear advantage on the $\log_e \propto$ scores ($p < .02$; 1-tailed Wilcoxon) and on percent correct ($p < .05$; 1-tailed Wilcoxon).

Discussion

This experiment certainly gives no support to the hypothesis that the competing stimulus must be part of the response set for a right-ear advantage to be obtained. Provided that the competing stimulus is from the same perceptual class, it need not be part of the response set. Thus a plosive can be an effective competing stimulus to another plosive even if it is not in the response set; by contrast, noise is not a sufficient competing stimulus. This result does not, of course, say at what level between these two extremes competition is effective. The result is quite compatible with the view that the ear difference effect is primarily a perceptual phenomenon but is not so readily explained by a view maintaining that only processes subsequent to phonetic categorization are pertinent.

The significantly greater performance on condition 2 (two sounds from the response set) than on either of the other two conditions suggests that a more predictable stimulus is more readily ignored than a less predictible one. This is, of course, confounded in the present experiment partly by the particular consonants used and partly by the voicing restriction in stimulus conditions 2 and 3. However, the effect is a large one, has implications for theories of attention, and warrants further research.

REFERENCES

Corsi, P.M. (1967) The effects of contralateral noise upon the perception and immediate recall of monaurally-presented verbal material. Unpublished M.A. thesis, McGill University, Montreal.

Darwin, C.J. (1971a) Ear differences in the recall of fricatives and vowels. Quart. J. Exp. Psychol. 23 46-62.

Darwin, C.J. (1971b) Dichotic forward and backward masking of speech and non-speech sounds. Paper presented to Acoust. Soc. of Amer., Washington, April 1971.

Haggard, M.P. (1968) Selectivity in speech perception models. Z. Phonetik 21, 70-73.

Haggard, M.P. (1971) Encoding and the REA for speech signals. Quart. J. Exp. Psychol. 23, 34-45.
Inglis, J. (1962) Dichotic stimulation, temporal lobe damage and the perception and storage of auditory stimuli: A note on Kimura's findings. Canad. J. Psychol. 16, 11.
Kimura, D. (1961a) Some effects of temporal lobe damage on auditory perception. Canad. J. Psychol. 15, 156-165.
Kimura, D. (1961b) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.
Kirstein, E. and Shankweiler, D.P. (1969) Selective listening for dichotically presented stop consonants and vowels. Paper read at Eastern Psych. Assoc. Meeting, Philadelphia, April 1969.
Luce, R.D. (1959) Individual Choice Behavior. New York: Wiley.
Mattingly, I.G. (1968) Experimental methods for speech synthesis by rule. IEEE Trans. AU-16, 198-202.
Milner, B., Taylor, L. and Sperry, R.W. (1968) Lateralized suppression of dichotically presented digits after commissural section in man. Science 161, 184-186.
Murphy, E.H. and Venables, P.H. (1970) Ear asymmetry in the threshold of fusion of two clicks: A signal detection analysis. Quart. J. Exp. Psychol. 22, 288-300.
Oxbury, S., Oxbury, J., and Gardiner, J. (1967) Laterality effects in dichotic listening. Nature 214, 742-743.
Sparks, R., and Geschwind, N. (1968) Dichotic listening in man after section of neocortical commissures. Cortex 4, 3-16.
Studdert-Kennedy, M. and Shankweiler, D.P. (1970) Hemispheric specialization for speech. J. Acoust. Soc. Amer. 48, 579-594.
Treisman, A.M. and Geffen, G. (1968) Selective attention and cerebral dominance in perceiving and responding to speech messages. Quart. J. Exp. Psychol. 20, 139-150.

The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables*

Robert J. Porter+
Haskins Laboratories, New Haven

The concern of this paper is the lag effect. This effect, first observed by Studdert-Kennedy, Shankweiler, and Schulman (1970), may be summarized as follows: if two stop consonant-vowel syllables are presented dichotically, with an onset asynchrony of 15 to 150 msec, subjects identify the temporally lagging consonant more accurately than the leading. That is, most errors are made in identifying the leading syllable of the dichotic pair. Dichotic presentation is essential for demonstrating the effect, for if the same temporally offset pairs are electronically mixed and presented to the same ear, then the leading syllable is identified more accurately than the lagging one. Studdert-Kennedy et al. suggested that the different effects for dichotic and monotic presentation might reflect the different influences of central and peripheral masking. The basis for their interpretation is diagrammed in Figure 1 for two syllables with an onset asynchrony of 75 msec. Each syllable is represented by a rectangle divided into two portions: the obliquely striped initial portion represents the location of the principle acoustic cues for the differing initial consonants; the final horizontally shaded areas represent the longer vowel portions which are the same for the two syllables. (The acoustic segmentation of consonant and vowel is intended for illustrative purposes only.)

Studdert-Kennedy et al. suggested that dichotically, at the top of the figure, the lagging syllable, which arrived centrally over a peripheral pathway separate from that of the leading, somehow disrupted the processing which had been initiated by the leading signal, information concerning the leading signal being lost as a result of this central disruption or "masking." Different factors were suggested as operating in the monotic case. Here, as is represented at the bottom of the figure, the two syllables physically overlap as they travel the same peripheral pathway. This overlap provides a sufficient condition for the peripheral masking of the lagging syllable's consonantal information by the simultaneously occurring overlapped vowel portion of the leading. Studdert-Kennedy et al. suggested that this peripheral masking of the lagging syllable by the leading overrode or precluded their central interaction and resulted in the observed advantage for the leading signal. Conceived in this way, the overlap of the signals is critically important for

---

*Paper presented at the 81st Meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

+Also University of Connecticut, Storrs.

Acknowledgment: This research was supported, in part, by grants to the author from NIH and the University of Connecticut Research Foundation.

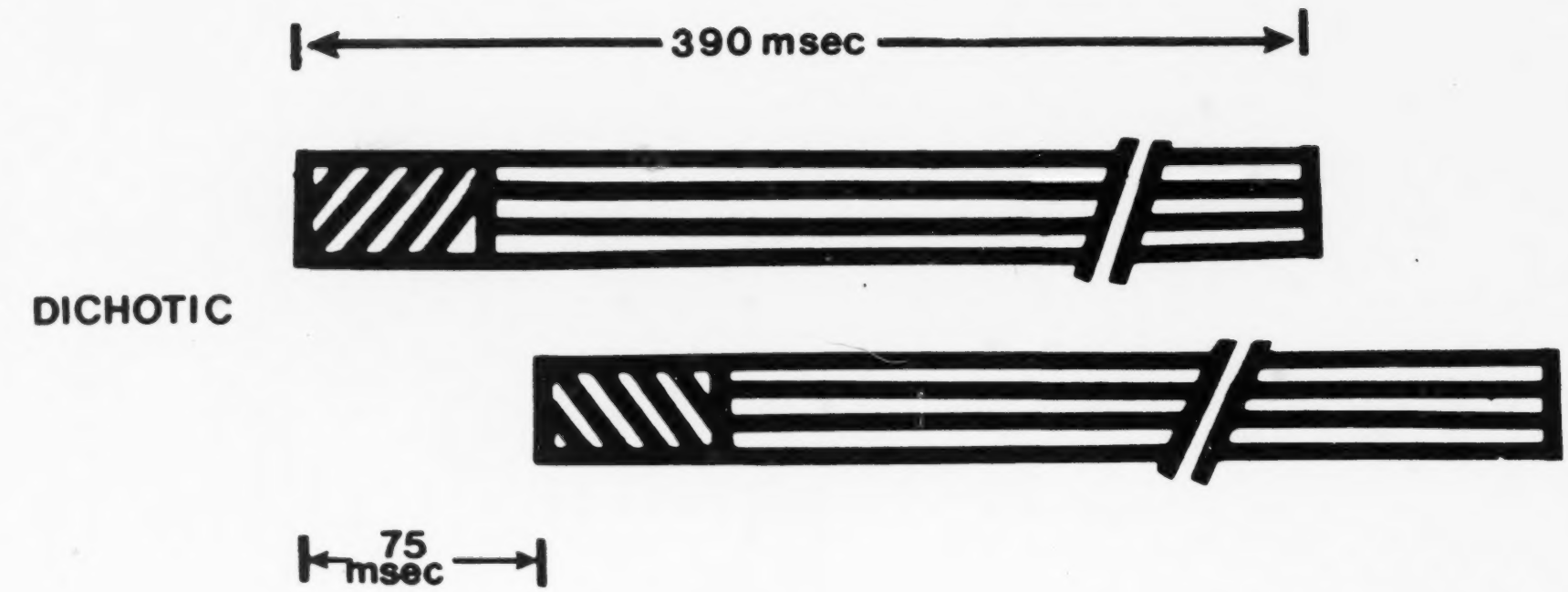


Fig. 1

demonstrating the monotic lead advantage. In the present study, the effect of eliminating the overlap was investigated by using syllables only 75 msec long and thus not overlapping at asynchronies of 75 msec or more.

Stimuli and tapes were prepared using the Haskins Laboratories computer-controlled parallel formant synthesizer. The stimuli were the six dichotic pairings of the stop-vowel syllables /ba, da, ga/. Six onset asynchronies of 0, 10, 25, 50, 75, and 100 msec were used. Two dichotic tapes were constructed: in one, the syllables were 390 msec long and thus overlapped at all asynchronies; for the second test, the syllables were 75 msec long and thus did not overlap at the two longest asynchronies of 75 and 100 msec. The 10-msec asynchrony was not included in the 75-msec syllable test. Corresponding long and short syllables were acoustically identical except for the duration of the final vowels. In each test, the dichotic pairs and the channel of the lagging syllable were appropriately counterbalanced with asynchronies. Six subjects (three males) received each dichotic tape twice on two successive days of testing. Following the dichotic tests on the second day, the subjects received each tape once with the two channels electronically mixed and presented binaurally. Pilot work had indicated that binaural presentation of the mixed channels did not yield results different from those obtained monaurally. For all tests, subjects were told to identify both consonants on each trial, guessing if necessary, and to record their responses on specially prepared answer sheets.

The results for all subjects combined are presented in Figure 2. The abcissa indicates the difference between the number of responses correct for the lagging and the number correct for the leading syllables expressed as a percent of the total number of syllables correct. Positive values indicate a lag advantage; negative values, an advantage for the leading syllable. The ordinate gives the onset asynchronies; the curve parameters are the experimental conditions.

Considering first the overall pattern of results, there is a clear separation of dichotic and binaural conditions. Dichotically, lag advantages are seen; binaurally, the leading syllable has the advantage.*

Dichotically, there is no significant interaction between length of syllable and degree of asynchrony. The scores for the shorter syllables appear to be somewhat lower than those for the longer syllables. This difference is not, however, systematically related to the progressive reduction in overlap with increasing asynchrony.

The decrease in amount of overlap has clear effects binaurally. There is a significant interaction between length of syllable and asynchrony. Lead

---

*The effects observed are generally smaller than those obtained in previous studies. This is probably due to the higher performance levels occasioned by the higher probability of guessing the other member of the pair given that one member is correctly perceived ($p = .5$). This is a consequence of there being only three tokens (as contrasted with six in previous studies). Subjects were aware of this fact and were told that the two syllables on any trial would be different.

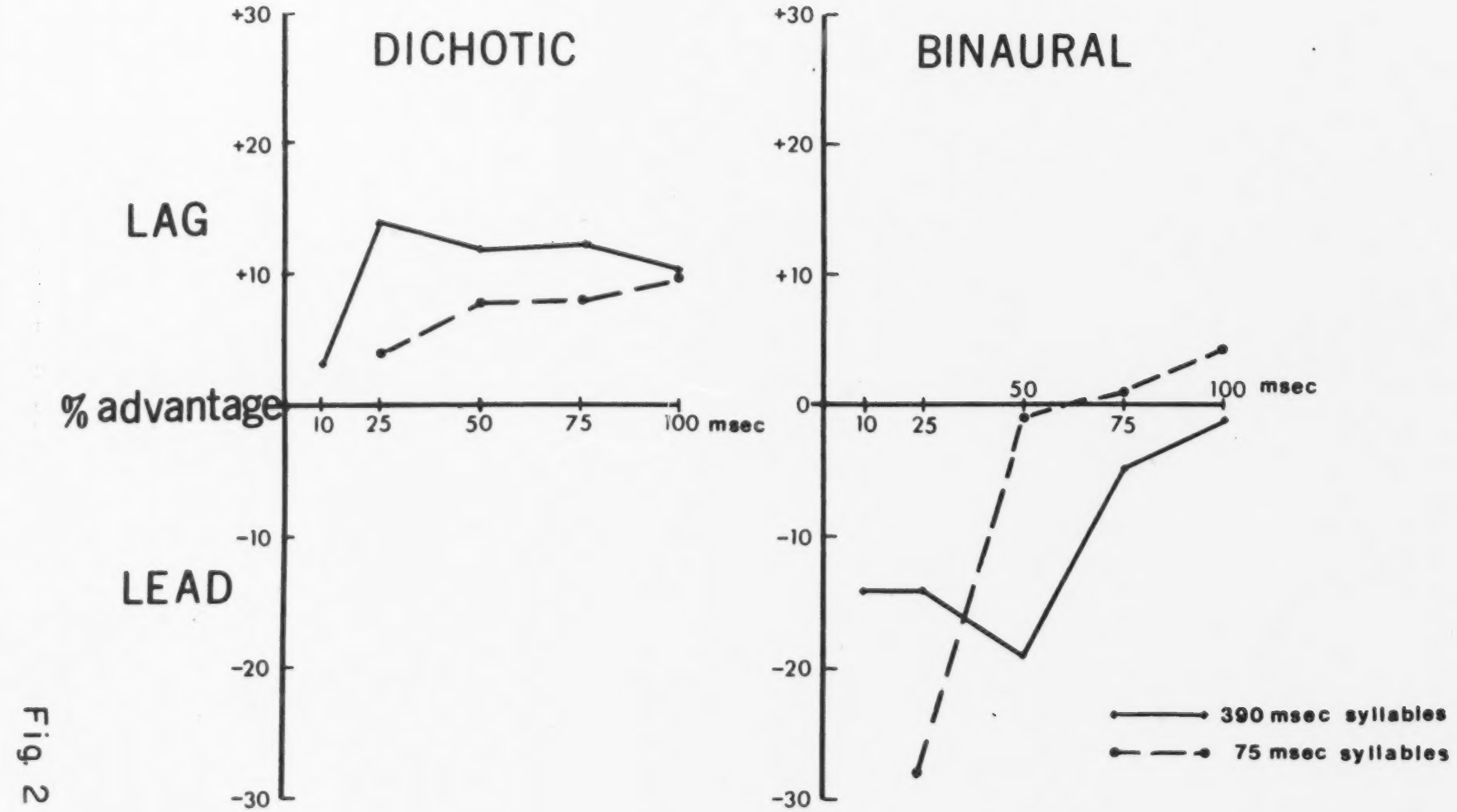

DICHOTIC
BINAURAL
+30
+20
+10
0
-10
-20
-30
LAG
LEAD
% advantage
10
25
50
75
100 msec
390 msec syllables
75 msec syllables

Fig. 2

advantages for the shorter syllables have essentially disappeared at asynchronies ≤ 50 msec (≤ 25-msec overlap), whereas the lead advantages for the longer syllables remain. A decrease in the amount of overlap would thus seem to reduce or eliminate the lead advantage seen binaurally, presumably because of the reduction in peripheral masking. These binaural results conform to subjects' postexperimental observations: with the longer overlapping syllables, only the leading one was heard clearly; with the shorter syllables, at 50, 75, and 100 msec asynchronies, both were heard. The results are perhaps not too surprising when we consider the fact that in rapid speech, consonantal information may often be temporally concatenated to nearly the degree that it is for these binaurally presented nonoverlapped short syllables.

The results do emphasize the interesting problem presented by the lag effect. Consider the short 75-msec syllables presented dichotically at 75- or 100-msec asynchrony. It is tempting to interpret the observed advantage for the lagging syllable in terms of a limited capacity of the central speech processor, that is, to suppose that the central processing initiated by the leading signal requires a certain amount of uninterrupted time in order to complete its necessary function. The untimely arrival of the lagging signal for some reason interrupts this processing, and as a consequence, information regarding the leading signal is lost. This account is not, of course, complete, since when the same temporal conditions are imposed binaurally, both syllables are processed with little difficulty. It is thus not simply the rapid temporal concatenation of information which overloads the central processor but rather the concatenation of information arriving over separate peripheral pathways. Work in progress is directed at determining why and in what way signal transmission over separate peripheral pathways places these unique constraints on central processing.

REFERENCE

Studdert-Kennedy, M., Shankweiler, D., and Schulman, Susan. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Auditory Evoked Potentials During Speech Perception*

Charles C. Wood,+ William R. Goff,++ and Ruth S. Day+++

ABSTRACT

Neural responses evoked by the same binaural speech signal were recorded from ten right-handed subjects during two auditory identification tasks. One task required analysis of acoustic parameters important for making a linguistic distinction, while the other task required analysis of an acoustic parameter which provides no linguistic information at the phoneme level. In the time interval between stimulus onset and the subjects' identification responses, evoked potentials from the two tasks were significantly different over the left hemisphere but identical over the right hemisphere. These results indicate that different neural events occur in the left hemisphere during analysis of linguistic versus nonlinguistic parameters of the same acoustic signal.

The relation between an acoustic speech signal and its phonetic message appears to be a complex and highly efficient code, which requires a specialized linguistic "decoder" for its perception (Liberman et al., 1967; Mattingly and Liberman, 1969; Studdert-Kennedy et al., 1970; Liberman, 1970). Dichotic listening experiments using normal (Kimura, 1961b, 1964, 1967; Shankweiler and Studdert-Kennedy, 1967; Curry, 1967; Curry and Rutherford, 1967; Kimura and Folb, 1968, Darwin, 1969a,b; Day and Cutting, 1970a,b; Studdert-Kennedy and Shankweiler, 1970) and brain-damaged subjects (Kimura, 1961a; Shankweiler 1966; Sparks and Geschwind, 1968; Milner et al., 1968; Schulhoff and Goodglass, 1969; Sparks et al., 1970) have further suggested that the specialized neural mechanisms required for the perception of speech are lateralized in one cerebral

---

*Paper to appear in Science.

+Neuropsychology Laboratory, Veterans Administration Hospital, West Haven, and Department of Psychology, Yale University, New Haven.

++Neuropsychology Laboratory, Veterans Administration Hospital, West Haven, and Departments of Neurology and Psychology, Yale University, New Haven.

+++Haskins Laboratories, New Haven, and Department of Psychology, Yale University, New Haven.

Acknowledgments: Supported by the Veterans Administration, U.S. Public Health Service NIMH Grant M-05286, and NSF Grants GB-3919 and GB-5782 to W.R. Goff, and by NICHD Grant HD-01994 to the Haskins Laboratories. We thank T. Allison, F.S. Cooper, M.W. Donald, D. Shankweiler, and M. Studdert-Kennedy for comments on the manuscript, and J.E. Cutting, T.C. Fisher, and J. Jasiorkowski for technical assistance.

hemisphere, usually the left. This interpretation is consistent with clinical analyses of language disorders following brain damage [1] and may be related to anatomical differences between left and right temporal lobes (Geschwind and Levitsky, 1968). In a recent review of hemispheric specialization for speech perception, Studdert-Kennedy and Shankweiler (1970, p. 579) concluded that "specialization of the dominant hemisphere in speech perception is due to its possession of a linguistic device....[W]hile the general auditory system common to both hemispheres is equipped to extract the auditory parameters of a speech signal, the dominant hemisphere may be specialized for the extraction of linguistic features from those parameters."

Despite the large body of behavioral and clinical evidence for specialization of one hemisphere in speech perception, there is no evidence which clearly distinguishes neural activity specifically related to linguistic processing from that which occurs during the processing of any auditory stimulus.[2] Empirical evidence for such a distinction requires a direct

---

[1] For a recent review see Geschwind (1970).

[2] Three experiments concerning neural activity evoked by speech sounds have been reported. Greenberg and Graham (1970) reported larger amplitudes of the evoked potential's "largest amplitude spectral component" from left- than right-hemisphere locations during a CV syllable learning task. No statistical evidence was included to show that the obtained results differed significantly from those expected by chance variation. Roth et al. (1970) reported no significant differences in activity recorded at the vertex to "sense and non-sense" monosyllables. In a paper published after the present experiment was submitted, Cohn (1971) reported "differential cerebral processing of noise and verbal stimuli." Cohn's major result was a "prominent positive-going peak with a latency of around 14 msec in the right brain derivation" in response to "click" stimuli generated by 10-msec pulses but not in response to "single syllable words" generated and presented in an unspecified manner. There are three major difficulties with the Cohn experiment: 1) No statistical evidence was presented to demonstrate that the obtained results differed significantly from chance variation. 2) Cohn's "verbal" and "noise" stimuli differed in many acoustic parameters such as duration, frequency composition, rise-time, total amplitude, and amplitude contour. Differences in neural activity evoked by such stimuli could be related to any or all of such acoustic differences, none of which need have any direct bearing upon the issue of speech versus nonspeech perception which Cohn wished to address. 3) It is possible that auditory evoked potentials of 14-msec latency are of nonneural origin. In previous reports of potentials recorded under conditions similar to those used by Cohn, the shortest latency potentials recordable from the human scalp and considered to be of neural origin do not occur until approximately 30 msec (Mast, 1965; Ruhm et al., 1967; Goff et al., 1969). Potentials in the 14-msec latency range have been considered to be nonneural artifacts (Bickford et al., 1964; Mast, 1965; Goff et al., 1969).

An experiment by MacAdam and Whitaker (1971) dealt with the question of hemispheric specialization of speech <u>production</u>. They reported slow potentials, distributed largest over the left hemisphere, occurring up to 1 sec before the production of polysyllabic words. Symmetrically distributed potentials were reported before the production of similar nonspeech gestures.

comparison of neural activity during linguistic and nonlinguistic processing conditions with other sources of variation in neural activity eliminated between conditions. We have therefore compared neural activity evoked by the same consonant-vowel syllable during two auditory identification tasks: one that required analysis of acoustic parameters which provide linguistic information (Stop Consonant Task) and one that required analysis of acoustic parameters which provide no linguistic information at the phoneme level (Fundamental Frequency Task). For convenience, we shall use the terms "linguistic and nonlinguistic parameters" to refer to those acoustic parameters that do and do not, respectively, provide linguistic information at the phoneme level.

Stop Consonant Task. Subjects were required to indicate which of two possible stimuli had occurred on each trial: /ba/ or /da/. The stimuli were generated by the Haskins Laboratories parallel resonance synthesizer and were prepared to be identical in duration (300 msec), initial fundamental frequency (F_0 = 104 Hz), frequency contour (falling), and intensity contour (falling). Thus, the two syllables differed only in those acoustic cues important for distinguishing between voiced stop consonants, namely the direction and extent of the second (Liberman et al., 1954; Delattre et al., 1955) and third (Harris et al., 1958) formant transitions. Stop consonants were selected for the linguistic task since they appear to be the most highly encoded of all phonemes (Liberman et al., 1967; Mattingly and Liberman, 1969; Studdert-Kennedy et al., 1970; Liberman, 1970).

Fundamental Frequency Task. Again subjects were required to indicate which of two possible stimuli had occurred on each trial. In this task, however, the two stimuli had identical linguistic information, namely formant transitions appropriate for the syllable /ba/. They differed only in fundamental frequency: /ba/-low (initial F_0 = 104 Hz) versus /ba/-high (initial F_0 = 140 Hz). Both stimuli were 300 msec in duration and had frequency and intensity contours matched to those of stimuli in the Stop Consonant Task. Variations in fundamental frequency were selected for the nonlinguistic task since absolute fundamental frequency provides little or no linguistic information at the phoneme level in English. Thus, the two tasks employed three acoustic stimuli, with the syllable /ba/-low (initial F_0 = 104 Hz) common to both tasks and used for comparison of evoked potentials. Spectrograms of the three stimuli are shown in Figure 1 arranged according to identification task.

Ten right-handed subjects (ages 18-20) were each tested during two separate sessions.[3] Both sessions consisted of six blocks of sixty-four stimuli, three blocks each of the Stop Consonant and Fundamental Frequency Tasks. A block of sixty-four stimuli contained thirty-two each of the two possible stimuli for that task, presented in random order at 5-sec interstimulus intervals. The two tasks were presented in alternating order during each session. Five subjects began session 1 with the Stop Consonant Task and session 2 with the Fundamental Frequency Task; the remaining five subjects

---

[3]Right-handed subjects were selected for this experiment since most are left-hemisphere dominant for language. See, for example, Zangwill (1960), Branch et al. (1964), Milner (1967), and Rossi and Rosadini (1967).

Spectrograms of the Three Stimuli

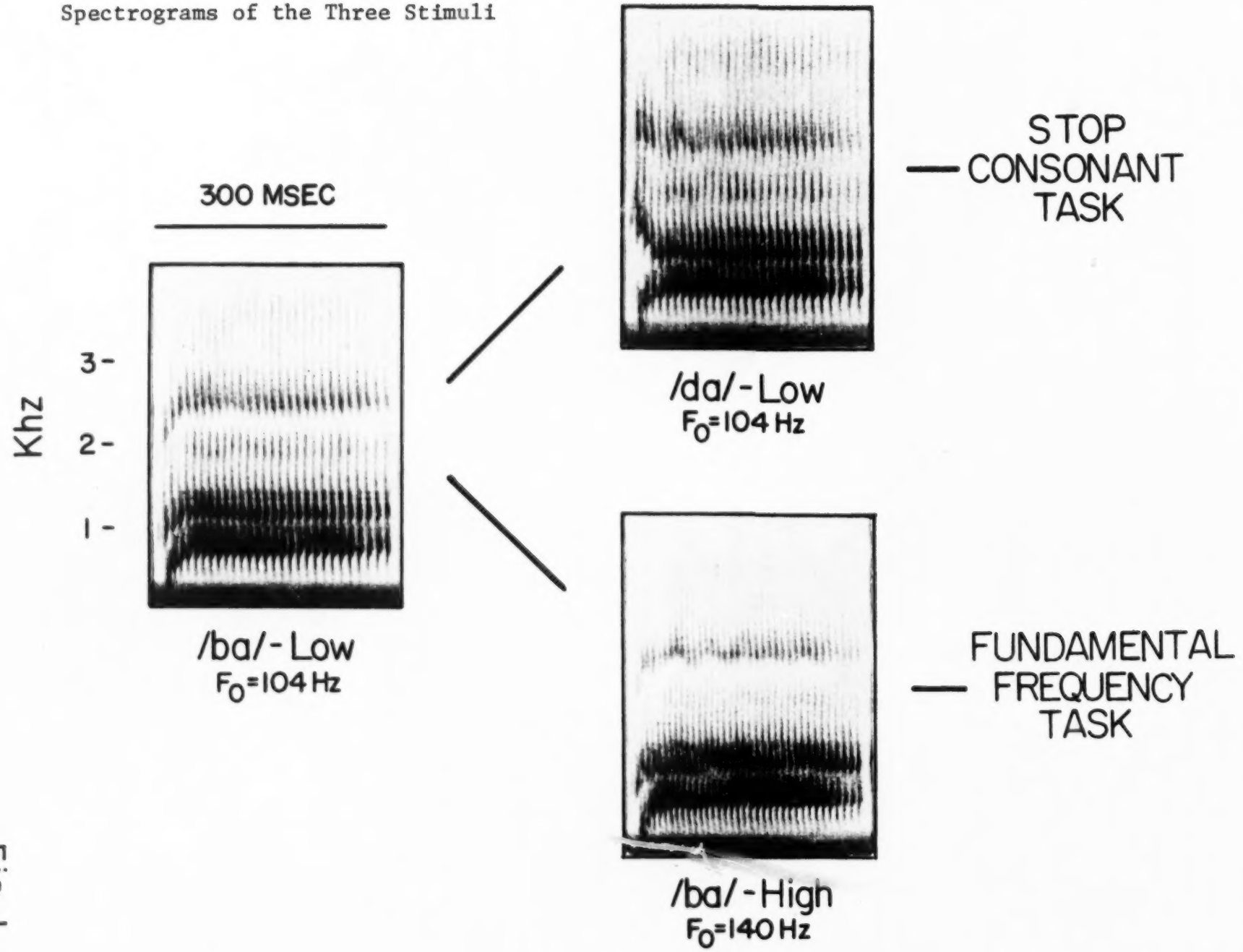


Fig. 1

Note: The syllable /ba/-low (F_0 = 104 Hz) was identical for both tasks. In the Stop Consonant Task, stimuli differed only in those acoustic cues important for distinguishing between voiced stop consonants (direction and extent of the second and third formant transitions). In the Fundamental Frequency Task, stimuli had identical linguistic information and differed only in fundamental frequency.

began the two sessions in the reverse order. Subjects were required to indicate which of the two possible stimuli they heard on each trial as soon as possible following stimulus onset. In both tasks, subjects pressed button 1 with the right index finger when they heard /ba/-low and button 2 with the right middle finger when they heard the other stimulus. Thus, both identification tasks contained an identical acoustic stimulus (/ba/-low), which occurred an equal number of times (thirty-two per run of sixty-four), with equal presentation probability on each trial ($p = .50$), and which required an identical motor response (pressing button 1 with the right index finger). Before session 1, subjects were asked to listen to the three acoustic stimuli and report what they heard. All subjects correctly identified each of the three syllables. They were then allowed to practice each task under conditions identical to those of the experiment until reaction times were stable. All subjects made fewer than five errors per run of sixty-four stimuli, and errors did not differ significantly between tasks. Therefore error scores will not be considered.

Electrical activity was recorded from temporal and central 10-20 system (Jasper, 1958) scalp locations over the left hemisphere (T3 and C3) and from corresponding locations over the right hemisphere (T4 and C4), each referred to a linked-ear reference using silver disc electrodes. Impedances of all electrodes were monitored regularly during each session and were less than 2.5 kilohms paired with the linked-ear reference. Particular care was taken to equalize impedances of the two ear reference electrodes: in all subjects both reference electrodes were equal at less than 3.0 kilohms, paired with each of the other electrodes.

Subjects were seated comfortably in a sound-attenuating and electrically shielded chamber illuminated at moderate intensity. EEG was recorded with a Grass Model 7 polygraph using Grass Model 7P5A wide-band A.C. EEG pre-amplifiers (system gain = 2×10^4) and was monitored visually throughout each run. Half-amplitude low- and high-frequency settings were 0.3 Hz and 500 Hz, respectively. Amplified signals were entered into a LINC computer for analog-to-digital conversion and signal averaging. Sampling epochs were 490 msec with 256 time points per epoch.[4] The LINC controlled the stimulus presentation order, averaged evoked potentials separately for each of the two stimuli in each task, and stored the averaged responses on magnetic tape for off-line data analysis. Subjects' identification responses and reaction times were recorded using a Beckman-Berkeley Model 7531R Universal Counter-Timer.

The synthetic stimuli were played to the subjects from a Precision Instrument FM tape recorder (Frequency response: $\pm$ 0.5 db, DC to 10 Khz at 30 ips). They were presented binaurally at 65 db SL against a 30-db white noise through a Grason-Stadler Model 829D electronic switch to G. C. Electronics earplug-type earphones. The timing of all events, including the initiation of LINC sampling epochs, was controlled by pulses on a separate channel of the FM tape recorder synchronized with stimulus onset.

---

[4]The 256 time points were distributed throughout the 490-msec epoch at three sampling rates: 1 point every 0.5 msec for the first 60 points, 1 point every 1 msec for the next 66 points, and 1 point every 3 msec for the remaining 130 points.

Averaged potentials evoked by the identical stimulus in both tasks (/ba/-low) were combined across subjects to obtain averages of 1,920 responses for each task and electrode location as shown in Figure 2. Evoked potentials from the Stop Consonant and Fundamental Frequency Tasks are superimposed at each electrode location to facilitate visual comparison. Reaction times did not differ significantly between tasks according to a Wilcoxon test (Siegel, 1956)(Median ± Semi-interquartile Range: Stop Consonant = 502 ± 75 msec, Fundamental Frequency = 493 ± 70 msec; T = 15, N = 10, $p > .10$). To determine the statistical reliability of differences between evoked potentials from the two tasks, Wilcoxon tests (Siegel, 1956) were computed between evoked potentials at each of the 256 individual time points in the sampling epoch.[5] Results of the statistical analyses are shown in Figure 2 below the evoked potentials at each of the four electrode locations. Upward deflections from baseline in the statistical traces indicate that the difference between evoked potentials at the time point was significant at the .01 level. For significance at the .01 level, the computation procedure for the Wilcoxon tests requires that the differences between evoked responses for a given time point occur in at least eight of the ten subjects.

In order to analyze evoked potentials during the identification processes required by the two tasks, the 490-msec evoked potential sampling epoch was empirically divided into the pre-response and motor response intervals shown in Figure 2. Since the identification process must be complete at or before the identification response is made, only the pre-response interval is appropriate for the analysis of evoked potentials during the identification process.[6] Differences between evoked potentials during the motor response interval will be considered below.

---

[5]This procedure was designed to determine a) the statistical reliability of differences between evoked potentials from the two tasks and b) the precise distribution of significant differences in time relative to stimulus onset and subjects' identification responses. Our procedure computed statistical significance for each of the 256 evoked response time sample points, using a standard non-parametric paired comparison technique (Wilcoxon matched-pairs signed-ranks test). At every sample point, the difference between amplitudes of responses from Stop Consonant and Fundamental Frequency Tasks was obtained for each of the ten subjects. The differences between tasks were then ranked and the Wilcoxon T statistic was calculated in the usual manner (Siegel, 1956). Thus, a value of the T statistic for the difference between evoked potentials from the two tasks was obtained for each of the 256 individual time points in each pair of responses.

[6]On a single trial the motor identification response unambiguously ends the time interval during which the identification process must have occurred. However, in the average of large numbers of trials required for comparison of evoked potentials, the proper end of the "processing interval" is less clear. Our criterion for distinguishing the pre-response and motor response intervals was the time point after which 99 percent of the motor responses occurred. The 99 percent point was selected instead of the 100 percent point because it disregards those few trials with extremely short RTs which cannot be meaningfully related to the identification tasks.

Electrical Activity Evoked by the Identical Stimulus in Both Tasks
/ba/-low (F_0 = 104 Hz)

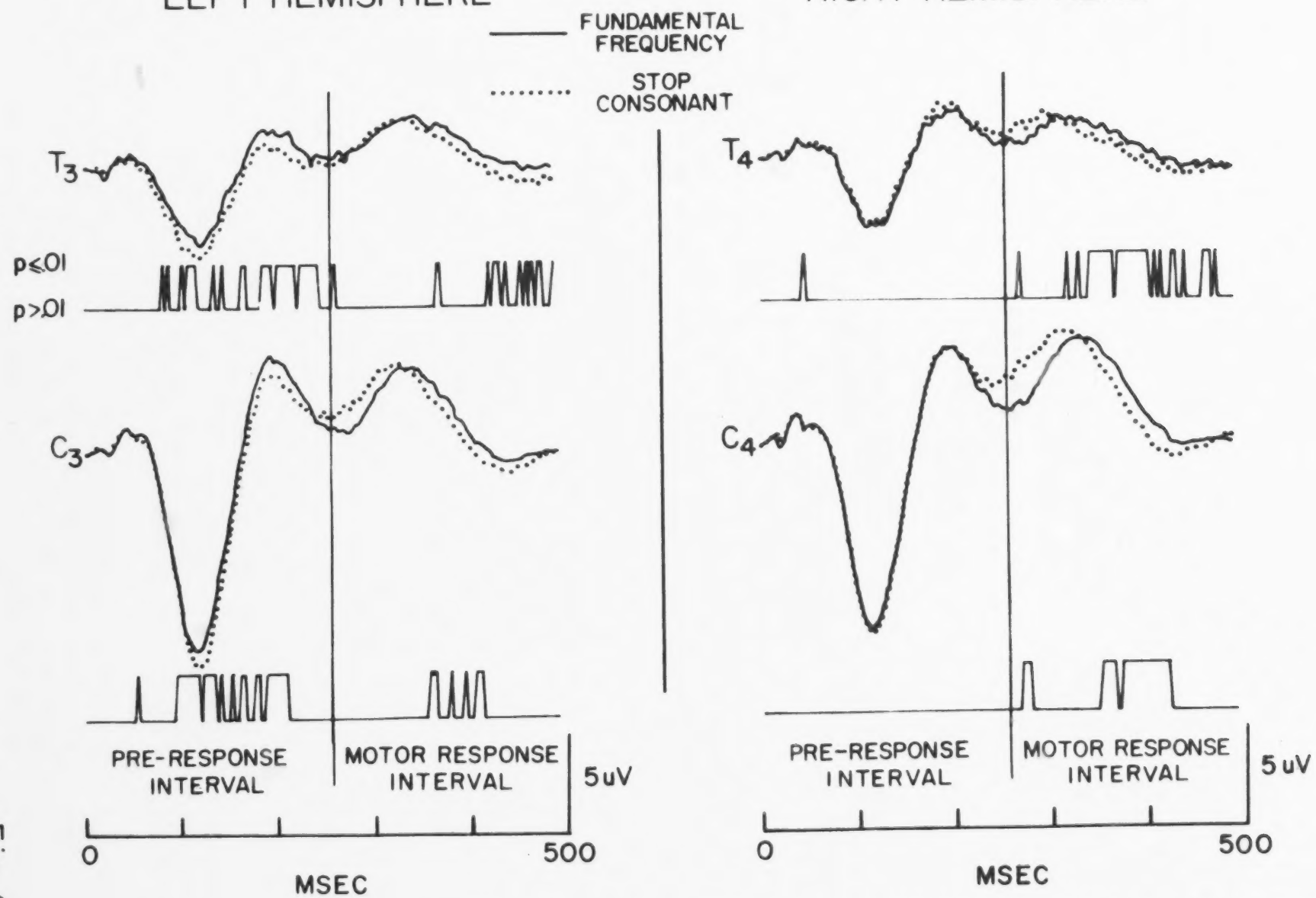


Fig. 2

Note: Averages of 1,920 responses for the Fundamental Frequency (solid lines) and Stop Consonant (dotted lines) Tasks are shown for left (T3 and C3) and right (T4 and C4) hemisphere locations. Positivity at the scalp electrode relative to the linked ear reference is upwards. Time scale: 500 msec. Vertical calibration: 5 uV. Results of the Wilcoxon statistical tests (Siegel, 1956) for each of the 256 individual sample points are shown below the evoked potentials at each electrode location. Upward deflections of the statistical traces indicate that the difference between evoked potentials at that individual time point is significant at the .01 level.

If the analysis of linguistic and nonlinguistic parameters of an acoustic signal consists of the same neural events, then evoked potentials should be the same (within the limits of normal variation) for both tasks during the pre-response interval. Evoked potentials from the right hemisphere (T4 and C4) were indeed identical for both tasks during pre-response interval, as shown in Figure 2. However, statistically significant differences in evoked potentials occurred at left-hemisphere locations (T3 and C3) during the same time interval. By chance variation, 1.77 significant time points would be expected at each location during the pre-response interval. At temporal and central locations over the left hemisphere 30 and 34 significant points were obtained, while 1 and 0 significant points were obtained at corresponding right-hemisphere locations. These results indicate that neural events in the right hemisphere were identical for both tasks during the pre-response interval, regardless of the task requirements. In contrast, different neural events occurred in the left hemisphere during the same time interval, depending upon whether the task required analysis of linguistic or nonlinguistic parameters of the acoustic signal.

We have been careful to eliminate factors which could produce artifactual differences in evoked potentials between tasks. There is, however, one additional source of possible artifact. Since the occurrence of a motor response (Karlin et al., 1970) and the speed of the that response (Bostock and Jarvis, 1970) can alter the neural activity evoked by sensory stimulation, it is possible that even nonsignificant differences in RT produced the results shown in Figure 2. To examine this possibility, the evoked potentials at each electrode location were recategorized. Instead of averaging the six Stop Consonant and six Fundamental Frequency blocks for each subject, the six fastest and six slowest RT blocks were averaged to maximize RT differences. Evoked potentials from the fast and slow RT blocks were then analyzed statistically in the same way as those in Figure 2.

If the evoked potential differences during the pre-response interval in Figure 2 were produced by nonsignificant differences in RT, then similar or larger differences should be produced by averaging the blocks with slowest and fastest RTs. Such a result did not occur. No more significant differences than would be expected by chance occurred at any electrode location during the pre-response interval: 1 significant point was obtained at each left-hemisphere location, and 1 and 2 significant points, respectively, were obtained at right-hemisphere locations. During the motor response interval, evoked potentials from the slow and fast RT blocks were significantly different in the same direction as those during the motor response interval in Figure 2. Thus, we cannot rule out the possibility that slight differences in RT may have produced the effects during the motor response interval shown in Figure 2. However, differences in RT could not have produced the significant differences in evoked potentials during the pre-response interval.

In summary, this experiment demonstrates that: 1) differences in neural responses evoked by the same speech signal occurred between tasks which required analysis of linguistic versus nonlinguistic parameters of that signal; 2) such differences occurred only at left-hemisphere locations; and 3) these differences are not related to differences in the acoustic signal, its presentation probability, the subjects' motor response, or reaction time. These results indicate that different neural events occur in the left hemisphere during analysis of linguistic versus nonlinguistic parameters of the same

acoustic signal. Further, they provide strong support for the idea that a unilateral neural mechanism is specialized to perform those linguistic processes necessary for speech perception.

REFERENCES

Bickford, R.G., J.L. Jacobson, and D.T. Cody. (1964) Nature of average evoked potentials to sound and other stimuli in man. Ann. N.Y. Acad. Sci. 112, 204-223.

Bostock, H. and M.J. Jarvis. (1970) Changes in the form of the cerebral evoked response related to the speed of simple reaction time. Electroenceph. Clin. Neurophysiol. 29, 137-145.

Branch, C., B. Milner, and T. Rasmussen. (1967) Intra-carotoid sodium amytal for the lateralization of cerebral speech dominance. Neurosurg. 21, 399-405.

Cohn, R. (1971) Differential cerebral processing of noise and verbal stimuli. Science 172, 599-601.

Curry, F.K.W. (1967) A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. Cortex 3, 343-352.

Curry, F.K.W. and D.R. Rutherford. (1967) Recognition and recall of dichotically presented verbal stimuli by right- and left-handed persons. Neuropsychologia 5, 119-126.

Darwin, C. (1969a) Laterality effects in the recall of steady-state and transient speech sounds. J. Acoust. Soc. Amer. 46, 114(A)

Darwin, C. (1969b) Auditory Perception and Cerebral Dominance. Unpublished Ph.D. thesis, University of Cambridge.

Day, R.S. and J.E. Cutting. (1970a) Levels of processing in speech perception. Paper presented to 10th meeting of the Psychonomic Society, San Antonio.

Day, R.S. and J.E. Cutting. (1970b) Perceptual competition between speech and nonspeech. Paper presented to 80th meeting of the Acoustical Society of America, Houston.

Delattre, P.C., A.M. Liberman, and F.S. Cooper. (1955) Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer. 27, 769-773.

Geschwind, N. (1970) The organization of language and the brain. Science 170, 940-944.

Geschwind, N. and W. Levitsky. (1968) Human brain: Left-right asymmetries in temporal speech region. Science 161, 186-187.

Goff, W.R., Y. Matsumiya, T. Allison, and G.D. Goff. (1969) Cross-modality comparisons of averaged evoked potentials. In Average Evoked Potentials, ed. by D. Donchin and D.B. Lindsley. (Washington, D.C.: NASA) 95-141.

Greenberg, H.J. and J.T. Graham. (1970) Electroencephalographic changes during learning of speech and nonspeech stimuli. J. Verb. Learn. Verb. Behav. 9, 274-281.

Harris, K.S., H.S. Hoffman, A.M. Liberman, P.C. Delattre, and F.S. Cooper. (1958) Effect of third-formant transitions on the perception of voiced stop consonants. J. Acoust. Soc. Amer. 30, 122-126.

Jasper, H. (1958) The ten-twenty electrode system of the International Federation. Electroenceph. Clin. Neurophysiol. 10, 371-375.

Karlin, L., M.J. Marty, and A.M. Mordkoff. (1970) Motor performance and sensory evoked potentials. Electroenceph. Clin. Neurophysiol. 28, 307-313.

Kimura, D. (1961a) Some effects of temporal lobe damage on auditory perception. Canad. J. Psychol. 15, 156-165.

Kimura, D. (1961b) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1964) Left-right differences in the perception of melodies. Quart. J. Exp. Psychol. 16, 355-358.
Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.
Kimura, D. and S. Folb. (1968) Neural processing of backwards-speech sounds. Science 161, 395-396.
Liberman, A.M. (1970) The grammars of speech and language. Cog. Psychol. 1, 301-323.
Liberman, A.M., P.C. Delattre, F.S. Cooper, and L.J. Gerstman. (1954) The role of consonant-vowel transitions in the perception of stop and nasal consonants. Psychol. Monogr. 68.
Liberman, A.M., F.S. Cooper, D. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.
MacAdam, D.W. and H.A. Whitaker. (1971) Language production: Electroencephalographic localization in the normal human brain. Science 172, 499-502.
Mast, T.E. (1965) Short-latency human evoked responses to clicks. J. Appl. Physiol. 20, 725-730.
Mattingly, I.G. and A.M. Liberman. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K.N. Lebovic. (New York: Springer) 97-116.
Milner, B. (1967) Brain mechanisms suggested by studies of temporal lobes. In Brain Mechanisms Underlying Speech and Language, ed. by C.H. Millikan and F.L. Darley. (New York: Grune & Stratton) 122-145.
Milner, B., L. Taylor, and R.W. Sperry. (1968) Lateralized suppression of dichotically presented digits after commissural section in man. Science 161, 184-185.
Roth, W.T., D.S. Kopell, and P.E. Bertozzi. (1970) The effect of attention on the average evoked response to speech sounds. Electroenceph. Clin. Neurophysiol. 29, 38-46.
Rossi, G.F. and G. Rosadini. (1967) Experimental analysis of cerebral dominance in man. In Brain Mechanisms Underlying Speech and Language, ed. by C.H. Millikan and F.L. Darley. (New York: Grune & Stratton) 167-184.
Ruhm, H.E. Walker, and H. Flanigin. (1967) Acoustically evoked potentials in man: Mediation of early components. Laryngoscope 77, 806-822.
Schuloff, C. and H. Goodglass. (1969) Dichotic listening, side of brain injury and cerebral dominance. Neuropsychologia 7, 149-160.
Shankweiler, D. (1966) Effects of temporal lobe damage on perception of dichotically presented melodies. J. Comp. Physiol. Psychol. 62, 115-119.
Shankweiler, D. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.
Siegel, S. (1956) Non-parametric Statistics. (New York: McGraw-Hill) 75-83.
Sparks, R. and N. Geschwind. (1968) Dichotic listening in man after section of neocortical commissures. Cortex 4, 3-16.
Sparks, R., H. Goodglass, and B. Nickel. (1970) Ipsilateral versus contralateral extinction in dichotic listening resulting from hemisphere lesions. Cortex 6, 249-260.
Studdert-Kennedy, M., A.M. Liberman, K.S. Harris, and F.S. Cooper. (1970) Motor theory of speech perception: A reply to Lane's critical review. Psychol. Rev. 77, 234-249.
Studdert-Kennedy, M., and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.
Zangwill, O. (1960) Cerebral Dominance and its Relation to Psychological Function. (Springfield: Thomas).

Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability*

Isabelle Y. Liberman+

ABSTRACT

Basic research in speech and the lateralization of language was shown to illuminate the problems of reading and some of its disabilities. First, it was pointed out how speech, or language for the ear, differs markedly from reading, or language for the eye. Though the sounds of speech are a very complex code and the optical shapes of written language are a simple cipher or alphabet on the phonemes, we all perceive speech easily but read only with difficulty. Perceiving speech is easy because, as members of the human race, we all have access to a special physiological apparatus that decodes the complex speech signal and recovers the segmentation of the linguistic message. Reading is hard because the phonemic segmentation, which is automatic and intuitive in the case of speech, must be made fully conscious and explicit. The syllabic method supplemented by phonics (used with certain reservations) was suggested for remediation of segmentation problems. Second, it was noted that since the sounds of speech are processed differently from nonspeech sounds, the two should not be diagnosed and remediated interchangeably. Third, it was shown that the relationships among cerebral lateralization for language, handedness, and poor reading can now be studied more meaningfully because of the recent development of new techniques.

A truism often heard in the opening lecture of graduate classes in education is that we have few answers to the problems that beset us, only questions. In the field of reading, the difficulty may be owing at least in part to our

---

*Article based on a talk given at the 21st Annual Meeting of the Orton Society, Washington, D.C., 14 November 1970. To appear in Annual Bulletin of the Orton Society, 1971.

+Department of Educational Psychology, University of Connecticut, Storrs. Though not a member of the staff of the Laboratories, the author has long been familiar with the work done here. This paper reflects her, and the Laboratories', interest in the implications of some of that work for an understanding of the reading process.

Acknowledgment: Much of the work reported here was done by researchers at the Haskins Laboratories. I should like to express my indebtedness to my husband, Alvin M. Liberman, and to a colleague, Donald P. Shankweiler, both of whom are members of the Haskins group and also professors of psychology at the University of Connecticut, for the assistance they have given me. Research sources other than the Haskins group are credited as they appear.

impatient attempts to find immediate solutions for the teacher and the student in the classroom and to our consequent neglect of basic research. I should like to suggest today how knowledge of basic research in related disciplines may lead to clues for improving beginning reading instruction and the lot of the disabled reader--if only by affording us a deeper understanding of the reading process.

THE POOR READER: DOES HE HAVE A "LANGUAGE DISABILITY"?

For over 75 years, much of the research in reading has been aimed at finding out how the poor reader differs from the good reader. Thus, many studies have correlated the reading level of the child with various indices of abilities or attributes which had been found to be defective in clinical studies of individual readers. These have, in the main, led to the conclusion that there are great individual differences among poor readers and that no single indices are typical of a large body of poor readers. The most consistent exceptions have involved tasks which are strongly language related, or actually reading related. I mean such tasks as oral word rhyming, oral vocabulary, word naming, letter naming, word recognition, name writing, and the like (De Hirsch, 1966; Doehring, 1968). Many, though not all, are essentially miniature reading and writing tasks. Of course, we should not need a giant correlational study to prove that reading is related to reading, nor should we be surprised to find that reading has something to do with language (though many remedial methods in current use seem to reflect this message only dimly).

It is certainly fair to say that in some sense the potentially poor reader frequently has language problems. But in what sense do we mean this? Given a child up to the age of eight--before his ability to read would make any substantial difference in his ability to speak--what is there about the language ability of the potentially poor reader that is different from that of the potentially good reader?

Data derived from two areas of basic language-related research seem to me to offer promising leads to these and other questions about reading--both the process and the disability. The two areas of basic research include speech perception and the lateralization of language.

I think we would all agree that poor readers can speak and listen to language far better than they can read or write it. From this point of view, to describe their problem as a "language disability" is to use the term very loosely indeed. Surely, if we could somehow teach them to write and read as well as they can speak and listen, we would not be concerned about their "language disability,"if any. Speaking and listening, then, are a necessary condition for reading but not a sufficient condition. It may be useful, therefore, to ask what we know about the difference between speaking and listening on the one hand, and reading and writing on the other.

LANGUAGE FOR THE EAR AND FOR THE EYE

We all know that human language is distinguished from other communication systems by the fact that it is phonemic. That is, all human languages are composed of commutable segments which have no meaning in themselves. It is clear that these phonemes can be transmitted either by ear or by eye--that

is, by spoken or written language.

Speech, but not Reading, is Natural to Man

We are all aware that speaking, or language for the ear, has a strong priority over reading, or language for the eye. The evidence for this is, of course, part of our common knowledge. Speech is universal, while reading is rare among the people of the world. Speech is first in the evolution of man, while reading is second; reading is, moreover, a comparatively recent development in man's history. It is also relevant to observe that the alphabetic method of reading and writing has been invented only once, which suggests that it is, in some important sense, unnatural. Speech is also first in the history of the individual while reading comes second. Speech is, moreover, remarkably easy for humans to acquire. Infants are already listening discriminatively to speech by the age of one month (Eimas et al., 1970) and most two-year-olds are beginning to speak intelligibly themselves. Speech apparently requires no tuition, only an input of linguistic data and an opportunity to interact with those data. In contrast, reading is difficult and is not ordinarily acquired unless it is taught.

The Sounds of Speech are Uniquely Natural

Moreover, as Mattingly and Liberman (1969) have pointed out, though sound is the only universal vehicle for the transmission of language, only one set of sounds, the sounds of speech, will work efficiently to transmit language. Morse code, which is an artificial sound alphabet, cannot be transmitted at rates much higher than five or six characters a second, even after years of practice. Other sound alphabets which were devised for use with reading machines for the blind seldom reached perceptual rates of more than two characters a second, though the subjects were often well practice and highly motivated. At rates far below those which are possible in the perception of natural speech sounds, the output of artificial sound systems become an unidentifiable blur to the perceiver. On the other hand, it is hardly necessary to remark that may alphabets--Cyrillic, Hebraic, Arabic, Roman, for example--are available and equally efficient for use in transmitting language for the eye, though none is as natural or easy as the sounds of speech.

The point I have been trying to make, then, is that speech and its sound are somehow basic to language in a way that the written language and its optical shapes are not. The phonemic segments of the language are transmitted easily and universally by the sounds of speech and by no others. Thus, the advantage is not with sounds in general, but very specifically with the sounds of speech. Optical shapes representative of language--the written letters of the alphabet--will also work to transmit the phonemic segments but they are a very recent invention in the history of man, are not used universally, and are relatively hard to use. With a few special and quite understandable exceptions, all human beings can speak and listen, but only a relatively few can read and, of that group, fewer still read well.

Transmission of Language by Speech Sounds and by Alphabetic Writing

We all know that speech and reading differ, as I have said they do, in the ease with which people master the processes. However, if our thinking has been conditioned by traditional views of speech perception and reading,

we may not have considered this to be a productive contrast to make. The traditional view includes two common assumptions about the transmission of language by ear and by eye which tend to obscure the important differences between these processes. Both of these assumptions are brought into serious question by recent research on speech.

The first false assumption is that the phonemic segments of language are transmitted individually by the sounds of speech, just as they are transmitted individually by the optical shapes of the alphabet. In this view, the sounds of speech bear a simple one-to-one relation to the phonemic segments, much as the optical shapes of the alphabet (orthographic variations aside) so obviously do. The word "bag," for example, which is represented in alphabetic writing by three letters, one for each of the perceived phonemic segments, is assumed to be represented similarly in speech by three discrete sounds. In this traditional view, then, whether the segments are represented by sound or by optical shape, the task for the perceiver would be basically the same, different only in that it is carried out in a different mode--in the auditory mode in the case of speaking and listening and in the visual mode in reading and writing.

Acoustic cues for the perception of speech. Let us see now in what ways this assumption may be false. Figure 1 shows at the top a speech spectrogram of the utterance, "Never kill a snake." A speech spectrogram is, of course, a visual display of the analyzed acoustic signal. Time is represented on the horizontal axis; frequency in cycles per second is represented on the vertical axis. The dark areas represent concentrations of acoustic energy at different frequencies for varying periods of time. As you can see, the spectrogram is a very "busy," muddy display.

People at Haskins Laboratories undertook to discover which aspects of this very complex signal carry the essential linguistic information. For this purpose, they developed techniques for converting spectrograms, including hand-painted versions, back into sound (Cooper, 1950, 1953; Cooper et al., 1951). Their aim was to find the more general nature of the relation between the acoustic signal, as seen in the spectrogram, and the phonemic message, which is what one perceives auditorally (Liberman et al., 1967; Mattingly and Liberman, 1969).

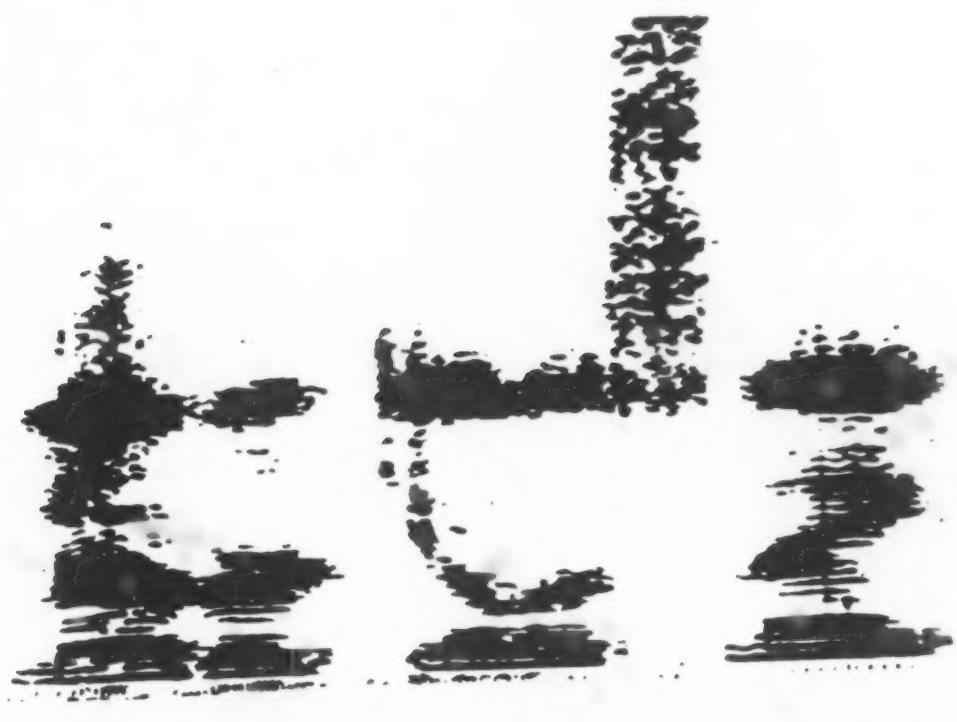
At the bottom of Figure 1 is a schematic painted spectrogram which represents a considerable simplification of the acoustical signal with the greater part of the signal discarded. The Haskins group found by trial and error that simplified spectrograms of this kind are nevertheless sufficient to produce intelligible speech. They proceeded, then, over a period of years to investigate this problem more systematically and succeeded in isolating the acoustic cues for all the various phonemic segements.[1]

Acoustic cues are not an alphabet on the linguistic message. Figure 2 shows examples of the essential acoustic cues for the universal stop consonant /d/ and also important general characteristics of the relation between the

---

[1]See Liberman et al. (1967) for a general review of these findings, together with references to the original experimental papers on which they are based.

A Spectrogram of the Phrase, "Never kill a snake"



A Simplified Hand-Painted Spectrogram Which is Sufficient, When Converted into Sound, to Produce an Intelligible Version of the Same Phrase

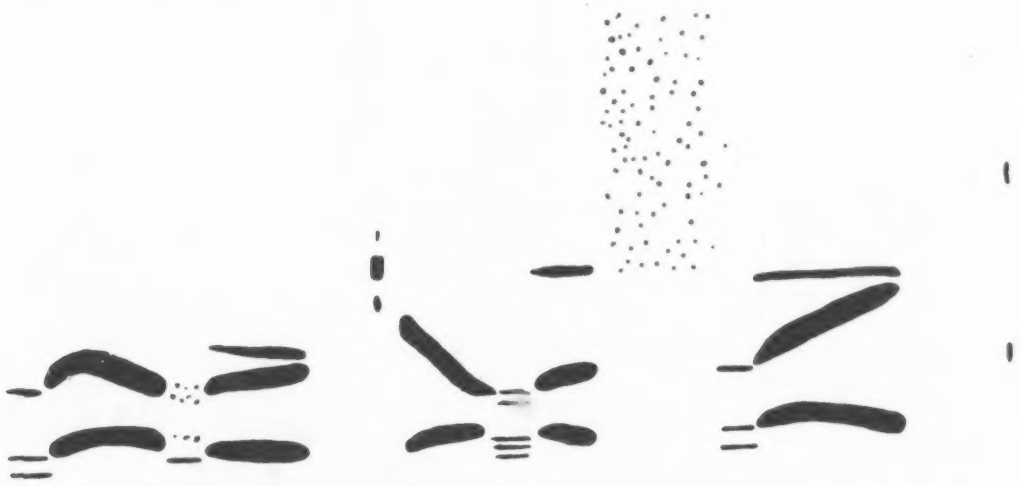


Fig. 1

Simplified Spectrographic Patterns Sufficient to Produce /di/ and /du/

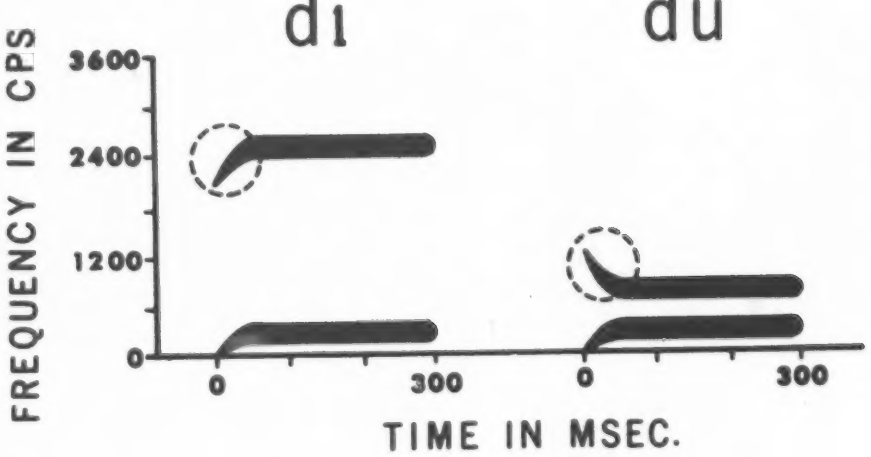


Fig. 2

(Reprinted from an article by A.M. Liberman in the October 1970 Cognitive Psychology. Copyrighted by Academic Press, October 1970.)

sound signal and the perceived message. The schematic patterns shown are sufficient for the synthesis of /d/ before /i/ and /u/. The black lines represent formants, i.e., concentrations of acoustic energy within a restricted frequency region. At the left of each pattern are the rapid changes in frequency known as the formant transitions. These have been found to be cues for the perception of consonants. The transition of the first, or lower, formant is the cue for the voiced stops--/b,d,g/. It carries the information about the manner and voicing of the consonant. This transition would be the same whether the syllables were /bi,bu/, /gi,gu/, or, as they are here in Figure 2, /di,du/, because /b,d,g/ are all voiced stop consonants.

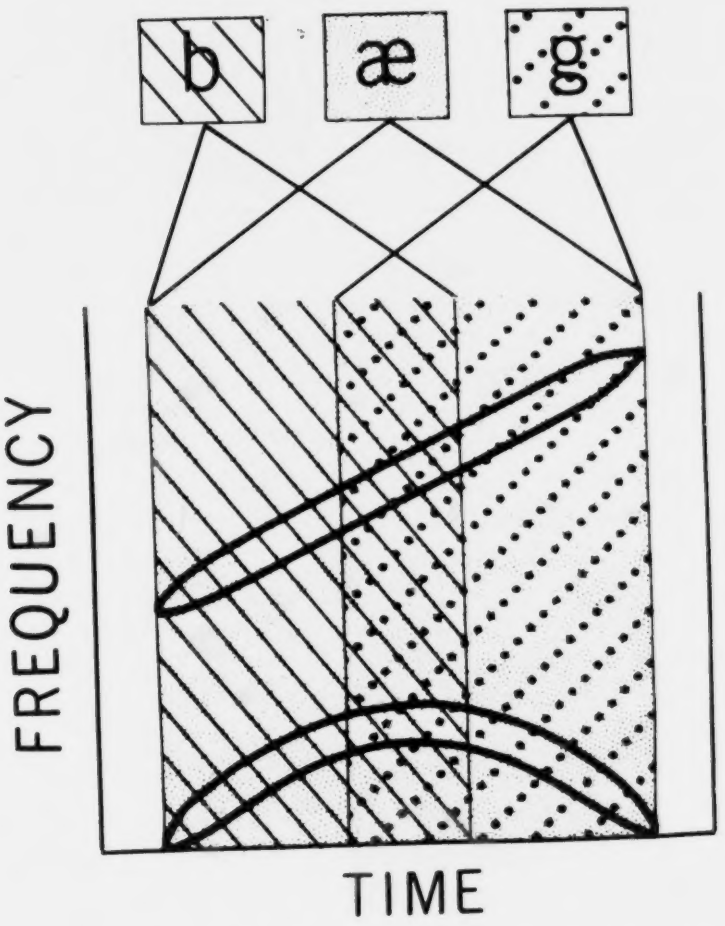
The second-formant transition, which is the part of the pattern circled in the upper formant, has been found to be the important acoustic cue for the perception of consonants according to their place of production. That is, in the case of stop consonants, it distinguishes /b/ from /d/ from /g/. In this figure, the second-formant transition contains the particular cue that causes the listener to hear /d/.

Now, in both syllables, /di/ and /du/, the /d/ sound heard by the listener is exactly the same. But the acoustic cues are very different in the two cases. In /di/, the second-formant transition rises from approximately 2200 cps to 2600 cps; in /du/, it falls instead from 1200 to 700 cps. Moreover, if one tries to separate these critical second-formant transitions from the context of the rest of the pattern and sound them in isolation, one does not get the /d/ sound at all. One gets nonspeech instead: a high-pitched rising whistle for /di/ and a low-pitched falling whistle for /du/. Outside the total pattern, the formant transitions sound very different from each other and neither of them sounds anything like /d/ (Liberman et al., 1967; Liberman, 1970).

We see, then, two related characteristics of the speech code: first, the acoustic cue for the same perceived consonant is different in two different vowel contexts, and second, there is no acoustic segment corresponding to the consonant segment /d/, for example. We cannot isolate the /d/ segment in the acoustic signal because the second-formant transition which is the essential cue for /d/ is always carrying information at the same time about both segments, the consonant and the vowel.

Successive segments of the message are complexly encoded in the acoustic signal. Figure 3 demonstrates more clearly how information about successive segments of the message is carried simultaneously by the same part of the speech signal. At the top are the perceived segments in the syllable "bag." At the bottom is a schematic spectrogram sufficient to produce that syllable. The figure shows how the segments which are experienced as separate at the perceptual level are intertwined in the sound stream. The vowel /æ/ is not limited to a medial position in the acoustic signal as it seems to be at the perceptual level but, rather, covers the entire length of the syllable. If the syllable were "big" instead of "bag," the second formant would be different from the beginning of the syllable to its end, not just in the middle position as it is in the perceived message. Similarly, information in the acoustic signal about the stop consonant /b/ continues well beyond the middle of the signal. If the syllable were "gag" instead of "bag," the second formant would change throughout the entire section subsumed under the segment /b/. Moreover, the center portion of the acoustic signal is obviously providing

Schematic Spectrogram Illustrating the Simultaneous Transmission of Successive Phonemic Segments on the Same Part of the Speech Signal



(Reprinted from an article by A.M. Liberman in the October 1970 Cognitive Psychology. Copyrighted by Academic Press, October 1970.)

Fig. 3

information not just about the vowel /æ/ but also about all three perceptual segments at once (Liberman, 1970).

All of this explains the failure of early investigators (Harris, 1953) to find the building blocks of real speech by cutting tape recordings into phonetic segments and then recombining the segments to produce new words. They could not do it, because, with one or two exceptions like steady-state vowels and parts of certain fricatives (Liberman et al., 1967), the perceived segments are not found as segments at the acoustic level at all.

Now we can get back to our original statement that the sounds of speech are not a simple alphabet or cipher on the phonemes as are the optical shapes of the written language. The sounds of speech are instead a very complex code. In this complex code, information about successive phonemic segments is transmitted simultaneously, not successively in strings as it is in the written language. For this reason, it is impossible to separate our discrete phonemic segments in any representation of the acoustic sound pattern.

THE POOR READER'S LANGUAGE PROBLEM: IS IT AUDITORY?

The Complex Speech Code is Handled Intuitively

When we consider again the child who speaks and listens so much better than he can read, we are faced with an interesting paradox. He can easily master the complex speech code and yet cannot master the relatively simple alphabet of written language. If speech does not appear complex to the human being who listens to it, it is presumably because he has ready access to the special neurophysiological apparatus necessary to handle it. There is now a great deal of evidence that such special processing equipment does exist as part of our human capacity for language. Later in this paper, I will describe just one aspect of that evidence. Meanwhile, we can observe that, as is the case with other biological processes that are deeply a part of us, we do not have to think about the process of speech in order to perform it, any more than we have to think about the process of walking in order to walk.

The Simple Alphabetic Cipher Requires Explicit Analysis of Language

If we now ask what the child is required to do in reading, we find a very different situation. There is, as we have said, a very simple relation between the alphabetic shapes and the linguistic message, but the child can take advantage of that relationship only if he explicitly analyzes and understands the segmentation of the message. Seeing the written word, being able to discriminate the individual optical shapes, being able to read the names of the three letters, and even knowing the individual sounds for the three letters, cannot help him in really reading the word "cat" (as opposed to memorizing its appearance as a sight word), unless he realizes that the word "cat" in his vocabulary has three segments. Before he can map the visual message to the word in his vocabulary, he had to be consciously aware that the word "cat" that he knows--an apparently unitary syllable--has three separate segments. His competence in speech production and speech perception is of no direct use to him here, because this competence enables him to achieve the segmentation without ever being consciously aware of it. (At the higher levels of language, similarly, one need not be consciously aware of the rules of grammar in order to produce grammatical speech.)

It seems reasonable, then, to suppose that the problem of this child who cannot read may not be, as is so commonly assumed, a problem in speech perception, or indeed, in auditory perception, at all. The intuitive and automatic segmentation he carries out in speech perception must be made quite conscious and explicit if he is to read; many children may find that extremely difficult. If so, what we are dealing with is a cognitive problem, not a problem of visual perception, auditory perception, or speech perception as such.

Implications of Speech Research for the Remediation of Reading Problems

The time-honored hypothesis offered when a child cannot understand that the components /k/ /æ/ /t/ form the word "cat" is that his difficulty lies in defective auditory perception or, more specifically, in not being able to blend sounds into words. One time-honored procedure for correcting this difficulty is to teach blending. I think you will agree with me that blending as either an explanatory or a remedial concept is now open to question. The word "cat" is not a blending of the sounds /k/ /æ/ /t/, if by blending one means a kind of merging of a string of consecutive sounds. It is clear that /k/ /æ/ /t/ merged together consecutively do not produce the word "cat." In speech, information about these three segments is encoded into a single sound, the syllable.

As might be expected, then, I would disagree with writers in the field (Johnson and Myklebust, 1967) who classify children with problems in phonetic analysis and synthesis as "auditory dyslexics" who have "numerous auditory discrimination and perceptual disorders which impede use of phonetic analysis [p. 174]." These writers themselves note that the spoken language of the children so classified "generally is good." I would say that, if the spoken language of the child is "generally good" and if he can respond appropriately to the speech of others, one cannot ascribe his difficulties with phonetic analysis and synthesis to poor auditory discrimination and perception. If he can hear and speak the words well, then his difficulty with segmentation is cognitive, not auditory.

Phonic, ideovisual, or syllabic method? I would agree that an elemental or phonic approach may be difficult for the child who cannot do phonetic analysis and synthesis, but I would strongly question the usual solution, which is to teach a sight vocabulary first--to teach by an ideovisual method, as it is called. If the child is indeed having difficulties with phonetic analysis and synthesis, then it would seem unwise to keep secret from him the relationship between the component parts of the spoken and written word. The sight method does just that when it proposes to teach the child to read by first teaching him to associate a certain whole spoken word with a particular whole printed design.

As I see it, it might be wise instead, to incorporate a type of syllabic approach into both beginning reading and remedial instruction. In this method, the component elements would not be treated separately as /k/ /æ/ /t/, but their identity would be clarified by the ordered use of phonetically regular syllables as suggested by Bloomfield (Bloomfield & Barnhart, 1961) and Fries (1962). By using the method of minimal contrasts and changing only one segment at a time in the syllables presented for study (e.g., "fat" and "mat," "fan" and "man"), one can illuminate the phonetic analysis of words from the start of reading instruction.

I am not prepared to say that analytic breakdown of words into their phonemes should not be used at all, but only that the ordered syllabic approach should also be used, because it conforms so much better to what we know about speech and language. And, moreover, if the phonic method is used, I would consider it important for the teacher to understand that when she uses the blending aspect of the instruction, she is not training the child's auditory perception of speech sounds. To the extent that she is helping the child at all, she is probably making it easier for him to achieve the conscious awareness of phonemic segmentation that he needs if he is to match the written version of the word to the spoken form already stored in his head.

Why vowels may present special problems. There is yet another result of speech research that may enlighten us about a difficulty commonly encountered in learning to read. A great deal has been made of the difficulties of the orthography, particularly in reference to vowel representation. There is, of course, no question that beginning readers find vowels more difficult to master than consonants. Every teacher can testify to this. Speech research indicates that there may be reasons for this that are not obvious on the surface. We learn from speech research that whereas consonants are distinctively categorical in both speech production and perception of speech, vowels are continuous and variable (Liberman et al., 1967). There is nothing between /b/ and /d/. There is only a /b/ and a /d/. When the acoustic cues for producing a /d/, for example, are changed in the direction of /b/, let us say, what you hear is either /b/ or /d/, never something in between. Consonants, then, and particularly the stop consonants (/b,d,g/ and /p,t,k/), are not regions lying along a continuum. They are categorical in the sense that they are either one consonant or another. Vowels, on the other hand, change continuously, like the pitch or loudness of tones. They do not fall into neat compartments the way most consonants do. Shankweiler has suggested (1967) that our tendency to perceive consonants categorically probably makes it easier for us to learn to associate them with graphic symbols. Similarly, the continuous nature of vowels may make it harder for us to learn their correspondences and may even account for their multiple spellings in the orthography of the language. Perhaps while consonants can best be taught by the syllabic method, vowels should be separated out for additional phonic study.

THE LATERALIZATION OF SPEECH AND NONSPEECH SOUNDS

Earlier I said that there are at least two false assumptions about speech which tend to confuse our thinking and reading. The first, which I have dealt with in the preceding sections, is that speech is a simple cipher on the phonetic message. The second, which I propose to discuss now, is that the process involved in the perception of speech sounds is the same as that involved in the perception of nonspeech sounds. The traditional view here is that all sounds are acted upon by the brain in much the same way, whether they are speech sounds or, say, household noises like the jangling of the doorbell or the crackling of paper. As I said before, one would expect, in view of the complex nature of the speech code, that we would need very special devices in order to process or decode it and that the mechanism of speech perception would be very different from that involved in the perception of other sounds. Some of the most compelling evidence which shows that the processing of speech sounds is indeed very special and quite different from that of nonspeech sounds comes from research in cerebral lateralization (the term "lateralization" here refers to the tendency of one side or hemisphere of the brain to take over certain functions).

Auditory Rivalry Technique Tests Cerebral Lateralization of Language

It has long been known that language disabilities of various kinds usually accompany injury to certain parts of the left cerebral hemisphere; injury to corresponding parts of the right hemisphere produces no such disruption of linguistic function. About ten years ago, a psychologist in Canada, Doreen Kimura, developed a bloodless, relatively simple, and potentially quite analytic method of studying lateralization of speech and nonspeech (Kimura, 1961). In her method, the investigator presents two different stimuli simultaneously to the two ears by means of stereo earphones. This "dichotic" presentation sets up a kind of rivalry between the two ears. When the subject is asked to report what he has heard, it is found that more stimuli are correctly identified from one ear than the other. Which ear wins out in the rivalry--that is, which one provides the greater number of correct answers--will depend on the kind of stimuli that have been used.

Many investigators have since found that, when the sounds presented are verbal, there is a right-ear advantage. This is true whether the stimuli are digits, words, or simple consonant-vowel nonsense syllables (Kimura, 1961; Shankweiler and Studdert-Kennedy, 1967). On the other hand, when the sounds presented are nonspeech sounds of any kind (melodies, environmental noises, sounds made by common objects, animal sounds, etc.), they all produce a left-ear advantage (Kimura, 1964, 1967; Knox and Kimura, 1970). Moreover, these effects are obtained in children as young as five years old, whether the method of report is verbal or nonverbal--that is, whether the child indicates what he has heard by repeating it verbally or by pointing to a picture of it or to the object itself (Knox and Kimura, 1970).

Speech Sounds and Nonspeech Sounds are Processed Differently in the Brain

The implications of these findings for the study of the lateralization of language are provided by current knowledge of the actions of the auditory pathways. While each ear has representation in both hemispheres, the contralateral representation is stronger than the ipsilateral (Rosenzweig, 1951). Moreover, there is evidence that when competing signals are presented to the two ears, the ipsilateral pathways are inhibited (Milner et al., 1968). Therefore, the interpretation of the right-ear advantage for speech and the left-ear advantage for nonspeech is that speech sounds require processing in the left hemisphere, while nonspeech sounds need to be processed in the right.

The fact that the sounds of speech are processed in one side of the brain and the sounds of nonspeech in the other strongly supports the assumption that they are processed in different ways. It is obvious that speech sounds must undergo some sort of auditory processing, of course--if an individual is deaf to sounds, he will not be able to hear speech--but it appears that the decoding of the complex speech code requires, in addition, physiological apparatus specialized for that purpose. It is also of interest that this apparatus is on the same side of the head as the apparatus which processes the syntactic and semantic aspects of language (Shankweiler and Studdert-Kennedy, 1967). This suggests again that speech is an integral part of language.

Remedial Implications of the Difference Between Speech and Nonspeech

The different processing required by the two kinds of sounds has practical implications for reading remediation. If one had strong evidence that a child

really did have deficits in the perception of speech sounds, one would not necessarily expect to improve his skills in perceiving speech by first giving him training in discrimination or identification of nonspeech noises, as is often done in remedial work. Sounds do not range on a simple continuum from simple environmental noises to speech. If the child is not required to respond to speech, he is not functioning in the speech mode and therefore is not using the processing required in speech. Speech processing goes beyond that required in the discrimination of nonspeech sounds and is carried on in a different part of the brain mechanism.

THE POOR READER: IS HE WEAKLY LATERALIZED?

We have said that speech and language are lateralized and that perception in the speech mode is primarily in the left hemisphere. To the extent that reading taps into the linguistic process, laterality may well be involved in reading as well. Why people who are lateralized well enough to speak and listen might not be lateralized well enough to read is not presently known. But weak cerebral lateralization has been implicated as a correlate of poor reading since the pioneering work of Orton in the thirties (Orton, 1937), who drew this conclusion from his clinical observations of the prevalence of uncertain handedness and ambidextrality among children with reading problems.

Two questions arise here. The first is whether children who cannot read well are indeed weakly lateralized for language. The other is whether handedness is an adequate indicator of brain lateralization for language. In Orton's time, and until recently, the two questions could not be separated. The only method readily available for judging lateralization for language was indirectly through such means as the testing of handedness and other peripheral preferences. Now, for research purposes, the auditory rivalry test provides a way of measuring brain lateralization for language more directly and with an independently validated technique (Branch et al., 1964). Studies using the auditory rivalry technique to explore the lateralization of children who are good and poor readers are as yet limited in number and inconclusive in results (Sparrow, 1968) but should in the future provide answers to the first question (I. Liberman et al., in progress).

As to the second question, concerning the use of handedness as an indicator of language lateralization, handedness has long been known to be related in some manner to language lateralization (Zangwill, 1960). However, we need to know more about the exact nature of the relationship, particularly in the case of self-classified left-handers and ambilaterals. In studying this relationship, one must take into account the fact that handedness is not an either/or proposition but, rather, a continuous variable (Benton et al., 1962; Annett, 1970) and the fact that the strength of handedness in various tasks is particularly variable in left-handers (Humphrey, 1951; Benton et al., 1962; Satz et al., 1967).

The relation between handedness and language lateralization has been studied in a doctoral dissertation recently completed at the University of Connecticut (Orlando, 1971), using left- and right-handed children as subjects. The results suggest that the relationship can be measured more meaningfully when both handedness and language lateralization (as measured by the auditory rivalry test) are regarded as continuous variables rather than as dichotomies. In addition, the study indicates that the relationship is strengthened when handedness is measured in terms of relative proficiency on manual tasks,

rather than in terms of manual preferences. Under these conditions, it is found that handedness and language lateralization are, in fact, strongly correlated, even in self-classified left-handers. Moreover, the results of the auditory rivalry test correlate more highly with the overall (joint) measure of handedness than does any single handedness task. This type of study has yet to be carried out in such a way that the results can be made to bear on the differences, if any, between the poor reader and the good reader, though some data, as yet unanalyzed, are already available (Shankweiler, et al., in progress).

SUMMARY

To summarize, I have tried to point out today how basic research in speech and language might illuminate some of the questions we have about reading and its disabilities. The first point was that speech is basic to language in a way that reading is not. We cannot have language without speech but we can and do have language without a written form that can be read. Speech is natural to us; reading and writing are not.

The second point I tried to make was that the sounds of speech are a very complex code and the optical shapes of the written language are a relatively simple alphabet on the phonemes, yet most of us have no difficulty with the speech code while many are unable to read. This is because we have special apparatus that enables us to deal easily and intuitively with language as received by the ear despite the great complexity of the process, but we need something more in the way of a conscious, cognitive analysis of the phoneme structure of language if we are to read. When a child has difficulty in reading because he cannot segment the words and syllables of his vocabulary into their constituent phonemic elements, the problem would seem to be a cognitive one, not a matter of visual or auditory perception.

The third major point I tried to make was that speech perception involves considerably more than auditory perception of nonspeech sounds. Speech sounds and nonspeech sounds are processed by different mechanisms in different parts of the brain and cannot be diagnosed or remediated interchangeably.

The lateralization of function in the brain brought me to the fourth point, the relation of language lateralization to reading disability, and its corollary, the relation of language lateralization to hand preferences and proficiency. Adaptations of a new method of measuring brain lateralization, the auditory rivalry test, promise to provide answers to the first question and have already afforded meaningful directions for further exploration of the record. Another productive new approach is to consider both handedness and brain lateralization for language as continuous rather than dichotomous variables.

My general message was that what is known from basic research in speech and laterality can lead to new hypotheses about the problems of the beginning reader and the poor reader. I hope you will agree that these kinds of research may bring us closer to solutions for these vexing problems than we have managed to come after so many years of product-oriented investigations.

REFERENCES

Annett, M. (1970) A classification of hand preference by association analysis. Brit. J. of Psychol. 3, 303-321.

Benton, A.L., Meyers, R.,and Polder, G.J. (1962) Some aspects of handedness. Psychiatrica Neurologia: Basel 144, 321-337.

Bloomfield, L. and Barnhart, C. (1961) Let's Read. Detroit: Wayne State University.

Branch, C., Milner, B.,and Rasmussen, T. (1964) Intercarotid sodium amytal for the lateralization of cerebral speech dominance. J. Neurosurg. 21, 399-405

Cooper, F.S. (1950) Spectrum analysis. J. Acoust. Soc. Amer. 22, 761-772.

Cooper, F.S. (1953) Some instrumental aids to research on speech. Report on the Fourth Annual Round Table Meeting on Linguistics and Language Teaching. Monograph Series No. 3. Washington: Georgetown University Press, 46-53.

Cooper, F.S., Liberman, A.M.,and Borst, J.M. (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proceedings of the National Academy of Sciences 37, 318-325.

De Hirsch, K. (1966) Predicting Reading Failure. New York: Harper & Row.

Doehring, D.G. (1968) Patterns of Impairment in Specific Reading Disability. Bloomington: Indiana University Press.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1970) Speech perception in infants. Science 171, 303-306.

Fries, C.C. (1962) Linguistics and Reading. New York: Holt, Reinhart & Winston.

Harris, C.M. (1953) A study of the building blocks in speech. J. Acoust. Soc. Amer. 25, 962-969.

Humphrey, M.E. (1951) Consistency of hand usage. Brit. J. Educational Psychol. 21, 214-225.

Johnson, D. and Myklebust, H. (1967) Learning Disabilities. New York: Grune & Stratton.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1964) Left-right differences in the perception of melodies. Quart. J. Exp. Psychol. 16, 355-358.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Knox, C. and Kimura, D. (1970) Cerebral processing of non-verbal sounds in boys and girls. Neuropsychologia 8, 227-237.

Liberman, A.M. (1970) The grammars of speech and language. Cog. Psychol. 1, 301-323.

Liberman, A.M., Cooper, F.S., Shankweiler, D., and Studdert-Kennedy, M. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Liberman, I., Shankweiler, D., and Orlando, C. (Research in progress) The lateralization of good and poor readers in the second and third grades.

Mattingly, I.G. and Liberman, A.M. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K.N. Leibovic. New York: Springer-Verlag, 97-117.

Milner, B., Taylor, L., and Sperry, R.A. (1968) Lateralized suppression of dichotically presented digits after commissural section in man. Science 161, 184-185.

64

Orlando, C. (1970) Relationships between language laterality and handedness in eight and ten year old boys. Unpublished doctoral dissertation, University of Connecticut.

Orton, S.T. (1937) Reading, Writing, and Speech Problems in Children. New York: Norton.

Rosenzweig, M.R. (1951) Representations of the two ears at the auditory cortex. Amer. J. Physiol. 167, 147-158.

Satz, P., Achenbach, K., and Fennell, E. (1967) Correlation between assessed manual laterality and predicted speech laterality in a normal population. Neuropsychologia 5, 295-310.

Shankweiler). (1967) Some observations on reading disability in relation to spee h. Paper presented at the annual meeting of the Academy of Aphasia, Ann Arbor, Michigan, October 1967.

Shankweiler, D. and Studdert-Kennedy, M. (1967) Identificatior of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.

Shankweiler, D., Liberman, I., and Orlando, C. (Research in progress) Handedness and language lateralization in poor readers.

Sparrow, S.S. (1968) Reading disability: A neurophysiological investigation. Unpublished doctoral dissertation, University of Florida.

Zangwill, O.L. (1960) Cerebral Dominance and its Relation to Psychological Function. London: Oliver and Boyd.

The EMG Data System

Diane K. Port
Haskins Laboratories, New Haven

During the last year and a half the entire EMG data collection and processing system used at Haskins Laboratories has been significantly improved. The use and insertion of wire electrodes is discussed elsewhere (Hirose, 1971). The present paper describes aspects of the data system for readers interested in a technical account of the system consistent with the suggestions of the "Report of the Committee on EMG Instrumentation" (Guld et al., 1970).

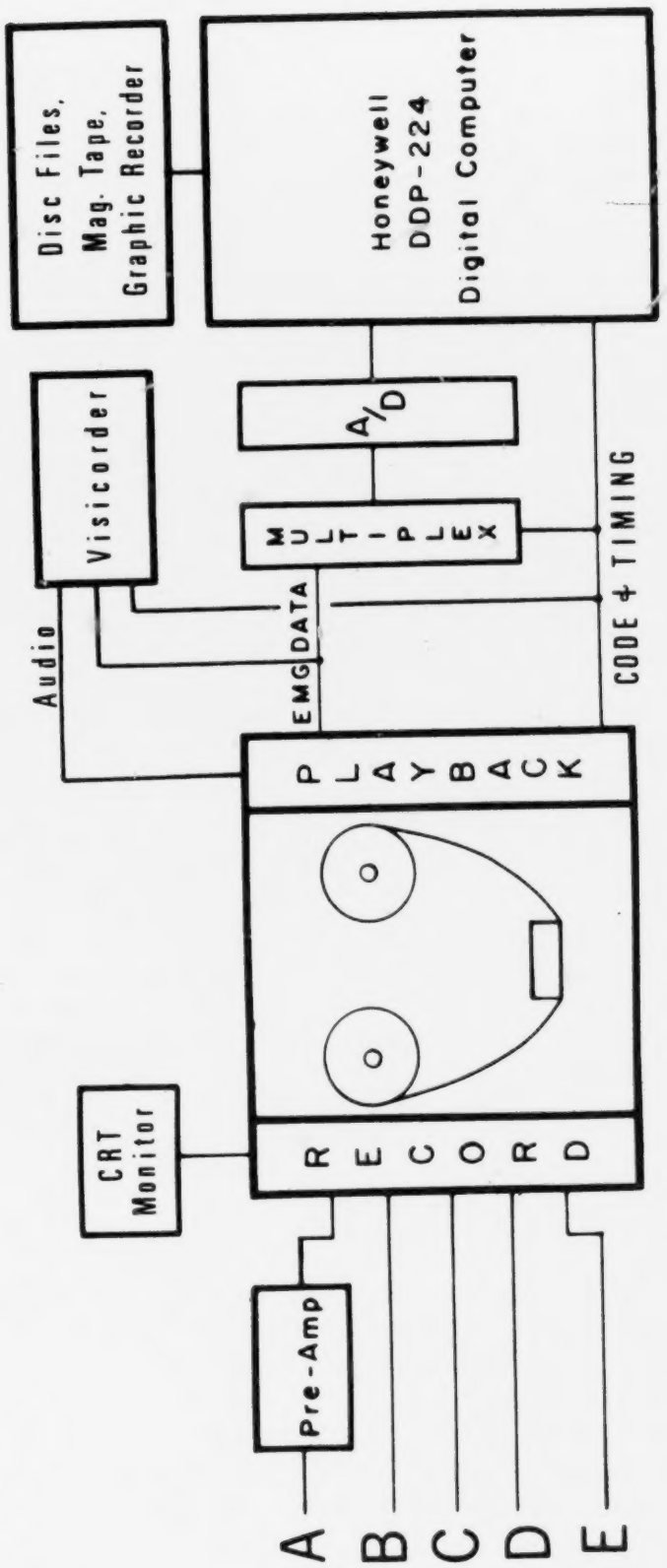
The basic principles of EMG data analysis used in earlier work at Haskins Laboratories have been carried over to the new system. Although the overall procedure for data collection and analysis has been described elsewhere (Cooper, 1965; Music et al., 1965; Sholes, 1965; Harris, 1970; Gay and Harris, 1971), a brief description will be repeated here for orientation.

An EMG experiment typically collects data on many repetitions of a limited set of utterances by one subject. The experiment proceeds in three principal stages: data collection, visual editing, and computer processing (measurement, averaging, and plotting). A major task in making this kind of experimentation feasible, i.e., in coping with the enormous amount of data involved in even a simple experiment, has been the development of procedures and equipment to automate most of the data collection and processing. The present system has largely accomplished this objective and has done so without sacrificing the experimenter's privilege of scrutinizing individual data entries to be sure that they are free of adventitious error. The equipment that has been assembled is shown diagrammatically in Figure 1. Its use will be discussed in the following descriptions of each of the three phases of a typical experiment.

DATA COLLECTION

The initial steps in data collection are, of course, the administration of such anesthesia as is required for patient comfort, insertion of the electrodes, and confirmation of their placement; this part of the procedure is described by Hirose (1971). The subject is then asked to read from randomized lists (or to repeat from a tape recording) the desired series of utterances, with pauses of a few seconds between them. Provision is made (in the programs that will later analyze the data) for up to thirty tokens of as many as thirty types of utterance.

The EMG signals collected by the bipolar wire electrodes go to differential preamplifiers which have gains of 40 db, noise levels (referred to the inputs) of 5 mv RMS, and ca. 100 db common mode rejection. From the preamplifiers, the signals go to distribution amplifiers with adjustable gains that are



A - EMG (8 Channels)

B - Air Pressure (2)

C - Voice

D - Banter

E - Digital Code & Timing

Fig. 1

67

usually set at about 30 db. These amplifiers include 80 Hz high-pass filters with 24 db roll-off to reject movement artifacts and hum. The filtered signals are then recorded on a one-inch, 14-channel instrumentation recorder (Consolidated Electrodynamics VR-3300). The EMG (and other physiological) data are recorded in FM; voice channels and timing and code pulses are recorded as AM signals.

A calibration signal (300 mv ± 1%) is substituted for each of the physiological signals several times in the course of an experiment. Periodic tests of the reference signal indicate that the long-term drift in the recording and amplifying equipment is less than 1% per month. The primary use of the calibration signals is, of course, to calculate the conversion of the physiological signals to microvolts at the electrodes.

Two recording channels are used for voice signals, one for the subject's utterances and the other for "banter" by the experimenters, in order to take note of events and changes in procedure during the course of the experiment. Two other channels are used to record a clock track and a code and timing track. The former consists of short pulses at a rate of 3200 Hz; the latter, of timing pulses at a rate of 50 Hz, counted down from the clock. Some of the pulses in the timing series are cancelled or inverted in polarity in order to generate a 4-digit octal code number that is incremented and recorded about once per second. (This way of introducing the identification codes has provided a good compromise solution to the problem of making the oscillographic record easily readable by humans and the tape-recorded version readable by computer.)

VISUAL EDITING

For visual inspection of the recorded physiological data, the data channels, voice channel, and code and timing track are played back as input to an 18-channel Honeywell Visicorder. During playbacks, the signals again go through the distribution amplifiers. Each physiological signal is routed through one section of its 80-Hz high-pass filter, resulting in a 36-db total roll-off. With the usual record/playback speed of 7.5 inches per second, the upper frequency limit of the FM channels is 1250 Hz. Thus, the overall frequency response (for EMG signals) is 80-1250 Hz. The signal-to-noise ratio for the FM channels is ca. 40 db.

The oscillographic traces for the voice and the code and timing marks are used by the experimenter to locate the specific portion of the EMG (and other) signals to be processed by the computer. This is done by first identifying each utterance with an octal code that precedes it. Then the temporal offset between this code and a distinctive event in the utterance (the line-up point) is noted. The choice of line-up point depends, of course, on the utterance; typical choices are stop-release or onset of voicing. The offset interval can be taken directly from timing pulses that occur at 20-msec intervals; typically, the offset interval is specified to the nearest 5 msec (a quarter of an interval on the timing trace), which is within the inherent uncertainty--estimated at ca. ± 10 msec--that is involved in locating the line-up point on the voice trace.

The two descriptors for each utterance (the octal code that identifies it and the offset interval between code and line-up point) are written down in lists by utterance type, and the lists are then entered into the computer.

These lists are merged, with the entries rearranged to be in the order in which the corresponding utterances appear on the instrumentation tape, and this merged list serves as the control information during computer processing.

COMPUTER PROCESSING

The measurement, processing, and plotting of the physiological data are almost completely automatic, although the experimenter can, if he wishes, intervene at various stages to test for, and correct, erroneous entries. The computer programs that make this possible were almost completely rewritten from the earlier programs--a necessary step in order to take advantage of substantial upgrading in the computer facility. Some of the capabilities that are important for EMG processing include four disc units (three for data) to allow for one-pass storage of all the digitized EMG signals for a complete experiment; magnetic tape for long-term storage (in digital form) of all the data generated in an experiment; and a strip chart recorder on which the final data (averaged for each electrode and utterance type) is plotted. All man-computer communication is through a Sanders Communicator, Model 720 (an alphanumeric CRT terminal). Programs for processing data are under control of a Monitor program (on the fourth disc unit) and several are currently being changed over to operate in a time-shared mode, to ease the requirements for computer time. Plans include a CRT display of any portion of the data for inspection and for automatic photography, if desired.

The programs are several in number and divide the processing task in the following way:

ESEL: Control information comprising the lists of codes and utterances already described is entered and stored in computer memory and on magnetic tape.

ECHK: The EMG signals are checked for correct control information, and analog input levels are set.

ERIT: The data are digitized and stored in one pass.

EDON: The signals are sorted and averaged, and the results are listed on a line printer.

E$MGPLOT: Hard-copy output curves are produced.

ESEL

The control information program is straightforward. Data about the utterances in the experiment and their line-up points are entered and stored on magnetic tape for later retrieval. Any item can easily be changed at any time during processing. The experiment size for which the ESEL program was designed is set at a maximum thirty lists of utterance types to be averaged, each of thirty speech utterances of 2-second maximum duration. Up to 8 channels of EMG data can be used.

ECHK

ECHK is the step in data processing that requires most operator attention. Here various checks on the offset between code and line-up point often catch gross measurement errors. There is also a print-out of the maxima and minima sampled in each channel for each code. By inspection of the consistency of these values for utterances of the same type, some obvious errors are detected. Errors are corrected in the control information before proceeding. Also at this time, the input gain levels (for playback from the instrumentation recorder) are set to make maximum use of the available digital data range.

ERIT

The digitizing program begins with playback of the instrumentation tape, the tape recorder being under computer control. The EMG signals in analog form are full-wave rectified and then passed through an RC circuit that performs a running integration. Typically, the time constant is set at 25 msec. The smoothed signals are sampled to 12-bit precision every 5 msec, using a 16-channel multiplexer driven by a clock that is internal to the computer but consistent with the recorded clock track to within 1%. Although twelve bits of data are delivered by the A-to-D converter and are recorded on disc and tape, only seven bits are significant since the system signal-to-noise ratio is approximately 40 db. Only the most significant seven bits are used later for averaging.

Given a 25-msec integration time constant on playback, a 5-msec sampling interval involves almost no loss of information due to sampling, according to the sampling theorem. Theoretical and empirical analyses of the effect of the RC integration circuit on the analog EMG signals is being undertaken and will be reported later.

EDON

In computing the EMG averages for each electrode location and utterance type, our first step is to convert the more-or-less arbitrary signal stored by ERIT to millivolts, using the recordings of the 300 mv reference signal that were made during data collection. Each of the reference signals specified in the control information is sampled and averaged over a 1-second interval. Then conversion factors are calculated for each channel and stored on magnetic tape. Next the sums and sums of squares for each utterance type (for each time sample from a given electrode location) are computed and stored on magnetic tape for further statistical analysis. Currently EDON calculates and converts to millivolts the means and standard deviations divided by the means. These values are printed out for the 5-msec intervals at which the analog data were sampled, referenced to the line-up point as time zero. It is possible, using EDON and ESEL to change any one of the utterance lists, for example, by deleting an erroneous code, and then to compute new sums and sums of squares.

E$MGPLOT

Hard copy is produced on a Texas Instruments Rectiriter Model RRMA strip chart recorder. The line-up point can be marked on the curves, if that is desired. The recorder is calibrated before each experiment.

SUMMARY

Overall, the new EMG data system appears to give a reliable output relative to actual myographic signals. The stability and signal-to-noise ratios of the system are good. Line-up points can be determined within ±10 msec. Gross errors are usually found and eliminated as a routine matter. The 25-msec time constant of integration is considered appropriate for the purpose of relating the high-frequency myographic signals to the comparatively slow movements of the articulators. (A time constant of 12.5 msec was tried and found to introduce more high-frequency noise without improving resolution of the averaged output curves.) Thus, the pattern of averaged outputs reflects mainly the pattern of muscle activity. Variability within an utterance type, as reflected in the standard deviation divided by the mean, shows token-to-token variation for EMG measures of the same utterance, though it does not necessarily imply as much variation in the muscle activity *per se*, since the EMG signal at each moment is determined by the relative phases of the signals from contributing muscle fibers, as well as by the total activity in the vicinity of the electrode.

REFERENCES

Cooper, F.S. (1965) Research techniques and instrumentation: EMG. *Proceedings of the Conference: Communicative Problems in the Cleft Palate*, ASHA Reports 1, 153-168.

Gay, T. and K.S. Harris. (1971) Some recent developments in the use of electromyography in speech research. JSHR 14, 241-246.

Guld, C., A. Rosenfalck, and R.C. Willison. (1970) Report of the Committee on EMG Instrumentation. Electroenceph. clin. Neurophysiol. 28, 399-413.

Harris, Katherine S. (1970) Physiological measures of speech movements: EMG and fiberoptic studies. ASHA Reports 5, 271-282.

Hirose, Hajime. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Music, R., L. Elbert, N. Reilly, and F.S. Cooper. (1965) Automating the collection and reduction of electromyographic data on speech production. Status Report on Speech Research (Haskins Laboratories) 1, 5.1-5.12.

Sholes, G.N. (1965) Measuring and averaging the data. Status Report on Speech Research (Haskins Laboratories) 1,6.1-6.9.

Electromyography of the Articulatory Muscles:
Current Instrumentation and Technique

Hajime Hirose*
Haskins Laboratories, New Haven

The particular merit of electromyography (EMG) in speech research is that it can provide information about the speech gesture in its natural units and that it directly reflects the motor command from the central nervous system carried by neural impulses. Recent technical developments in EMG have made possible examination of the articulatory muscles without affecting natural speech performance. The present report describes a current technique used at Haskins Laboratories for the assessment of EMG data from the human articulatory muscles.

ELECTRODES

At present, hooked-wire electrodes are used exclusively. The wire currently in use is a platinum-iridium alloy (90%-10%) with polyester (Isonel) coating, the diameter of which is 0.002 in. (Consolidated Reactive Metals, P-91). This wire is ideal for these experimental purposes, since it is less easily crimped or bent than copper wire and less springy than stainless steel In addition, the wire is possibly less irritative to human tissues than are other kinds of metal since no chemical reaction is to be expected.

The electrodes are made in essentially the same way as described by previous authors (Hirano and Ohala, 1969; Basmajian and Stecko, 1962). After a wire that is long enough to serve as a pair of electrodes (50-60 cm for percutaneous insertion and 80-90 cm for peroral insertion) has been prepared, the two free ends are threaded into the tip of a hypodermic needle (26 or 27 gauge and 3/4 to 2 in. in length) and pulled through the needle until a small loop remains. The loop is bent and cut with a razor blade to leave two short hooks of approximately 1-2 mm at the tip of the needle. Care is taken to make the two hooks of different lengths so as to avoid a possible short circuit by contact of the two cut ends in the muscle. The other ends of the wire are burned in a match flame to remove the polyester coating for connection to a preamplifier of the recording system.

For peroral insertion into the velopharyngeal muscles, the shaft of an electrode-bearing needle can be angulated to allow easier access to the target muscles.

For EMG of some laryngeal muscles, a specially designed probe is used for peroral insertion by indirect laryngoscopy. The probe consists of an L-shaped metal rod and the shaft of a 26-gauge needle cut and epoxy-bonded to the end of the shorter arm of the rod. The hooked-wire electrodes are made by threading the wire through the carrier needle in the conventional manner and also through a thin polyethylene tube bonded along the rod (Figure 1).

*Also Faculty of Medicine, University of Tokyo.

An L-shaped Probe Used for Peroral Insertion of the Hooked-Wire Electrodes into the Posterior Cricoarytenoid and the Interarytenoid

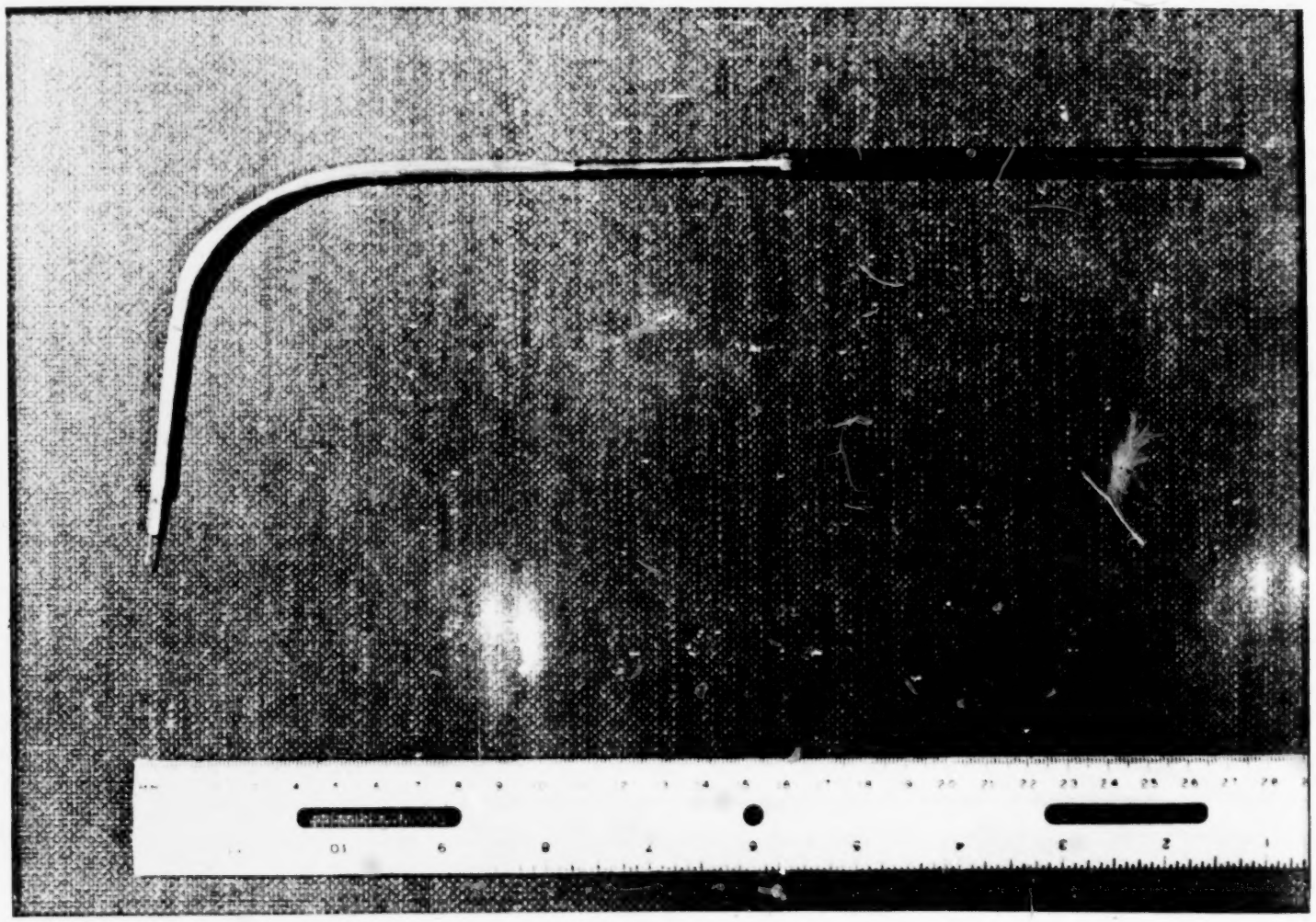


Fig. 1

Note: The shaft of a 26-gauge hypodermic needle is epoxy-bonded to the shorter arm.

The advantages and disadvantages of hooked-wire electrodes have been discussed so fully by previous authors (Hirano and Ohala, 1969; Harris, 1970) that no further comment will be made in the present report.

GENERAL PROCEDURES

Sterilization of the needle and the wire electrodes is accomplished either by high-pressure heat or by antiseptic solutions.

Before each experiment, if it is deemed necessary to inhibit salivation, 7-10 drops of Tincture of Belladonna is administered by mouth. For peroral insertion of the electrodes, topical anesthesia is administered to the pharynx using Cetacaine[1] spray and to the larynx, by this same method, in the case of laryngeal EMG. This is followed by a gargle or instillation of 2-3 ml of 2% Xylocaine.[2] The percutaneous insertions are preceded by topical administration of 2% Xylocaine without epinephrin through a Panjet-70 air jet (Panray)[3] at the site of the needle insertion.

The skin is disinfected at the site of insertion with an alcohol swab. A ground electrode (a gold earring) is attached to the ear lobe of the subject. During electrode insertion, an oscilloscope and an amplifier-speaker system are used for monitoring the pertinent muscle activity. After insertion into an appropriate site, the electrode-bearing needle is withdrawn leaving the electrodes hooked in the target muscle.

Whatever position is taken during electrode placement and its verification, the data recording is made with the subject in an upright sitting position. Oscilloscopic monitoring of selected EMG channels is provided throughout the procedure.

INSERTION TECHNIQUES AND VERIFICATION OF ELECTRODE PLACEMENT

Correct placement of the electrodes in the target muscle is prerequisite to the entire experimental procedure in an EMG study. The exact placement of the electrodes is easier if (1) the target muscle is close to or immediately beneath the covering skin or the mucosa and the insertion is possible under direct inspection or (2) there is little possibility of contamination with other muscles. In any case, verification of electrode placement is absolutely necessary.

---

[1] Cetacaine (trade name) is packaged in a 50-ml aerosol bottle and contains the following: ethyl aminobenzoate, 14%; butylamino-benzoate, 2%; benzalkonium chloride, 0.5%; cetyldimethylethyl ammonium bromide, 0.005%. A one-second spray releases 0.1 ml of solution, and usually three to four seconds of spray are needed to anesthetize the oral and pharyngeal mucosa. (Gaskil and Gillies, 1966).

[2] Recent studies (Shipp, 1968; Zemlin, 1969) revealed no discernible effect of topical anesthesia on normal laryngeal behavior.

[3] Panjet-70 delivers approximately 0.1 ml of the anesthetic solution to circumscribed intradermal depth up to 6 mm penetration.

In principle, correct placement of the electrodes is verified by monitoring the muscle activity induced by appropriate gestures that have been considered pertinent for the contraction of the target muscle. For some articulatory muscles, however, there is a lack of normative EMG data on which our verification can rely, as Shipp et al. (1968) have pointed out. Verification thus depends to a certain extent on the experimenter's empirical judgment based on his knowledge of anatomy and clinical and experimental practice. Further effort will be needed to reach unanimous agreement on the normative behavior of the articulatory muscles, although some interindividual variation both in anatomy and in function must always be taken into consideration.

Instrinsic laryngeal muscles

Posterior cricoarytenoid (PCA). The PCA is reached perorally by indirect laryngoscopy using the L-shaped needle holder described above. Using this approach, one can insert the needle parallel to the alignment of the muscle fibers; insertion is performed under inspection with a laryngeal mirror. During insertion, the subject is in a sitting position and is asked to phonate a sustained vowel so as to open his hypopharyngeal lumen for easier access to the site of insertion, which is illustrated in Figure 2. The insertion is thus made into the belly of the muscle on the cricoid cartilage through the hypopharyngeal mucosa. By this approach, there is little possibility of contamination with neighboring muscles unless the insertion is made too cranially.

Identification is made by having the subject repeat short periods of vowel phonation interspersed with deep inspiration. The PCA is active for inspiration and suppressed for the period of phonation, and this pattern is very characteristic.

Interarytenoid (INT). The insertion into the INT can be made either perorally or percutaneously. For peroral approach, the same technique is used as for the PCA and insertion is made at the midline between the two arytenoid prominences (Figures 3 and 4). By the percutaneous route, the transporting needle is inserted through the cricothyroid space, penetrating the skin and the cricothyroid membrane at the midline. Under inspection with a laryngeal mirror, the needle is pushed backwards and slightly upwards so as to pierce the anterior wall of the interarytenoid region to reach the INT (Figures 3 and 4). Both approaches are made with the subject in a sitting position. There is almost no possibility of contamination with other muscles in this case.

Verification of the placement is made by asking the subject to repeat short periods of phonation. The general pattern of INT activity is almost reciprocal to PCA; there is marked activity for the period of phonation.

Cricothyroid (CT). The percutanous route is always taken with the subject in a supine position. Insertion is made at a point above the cricoid ring and approximately 1 cm lateral to the midline. The needle is directed posterolaterally and slightly upwards aiming at the lower edge of the thyroid lamina. This is the same technique as reported by Hirano and Ohala (1969).

Verification of the correct placement is made by asking the subject to attempt an ascending scale. The CT shows marked activity for a quick rise in fundamental frequency.

A Diagrammatic View of the Larynx During Sustained Phonation by Indirect Laryngoscopy

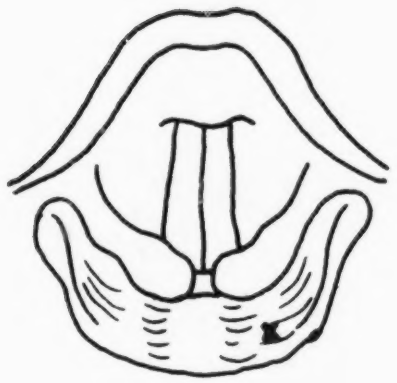
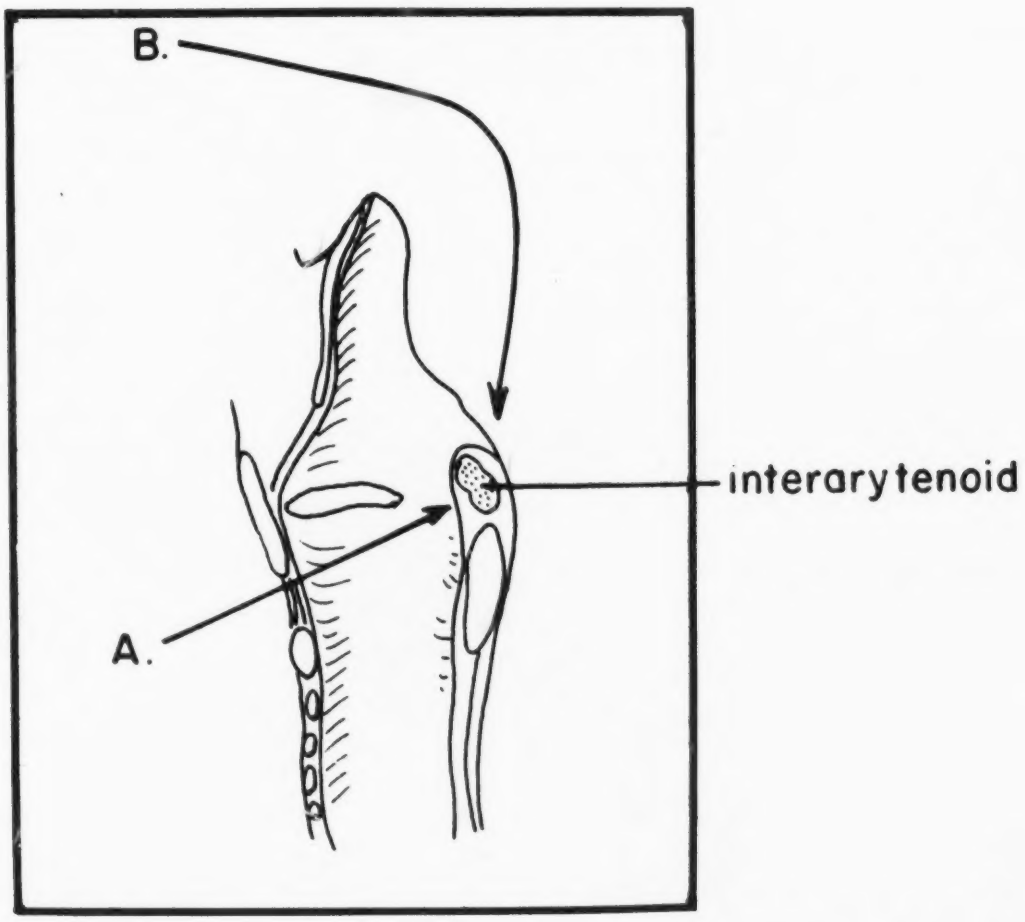


Fig. 2

Note: A cross (x) indicates one point of needle insertion into the posterior cricoarytenoid

A Sagittal Section of the Larynx with Illustration of the Direction of Needle Insertion to the Interarytenoid



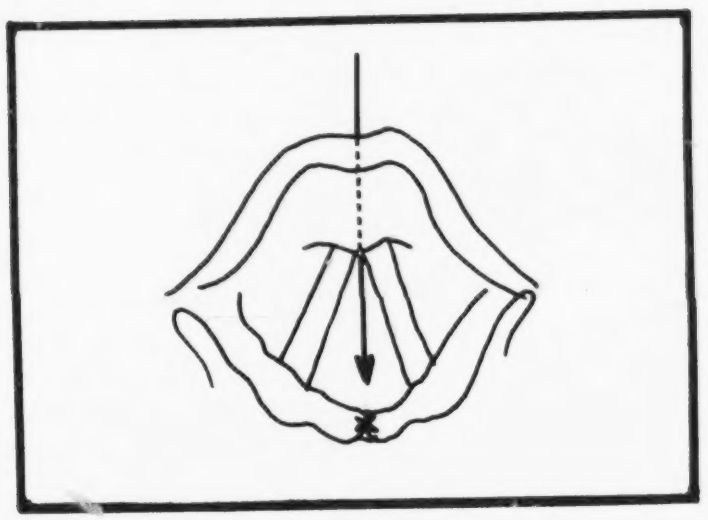
A. Percutaneous Route

B. Peroral Route

Fig. 3

77

Indirect Laryngoscopic View of the Larynx During Quiet Respiration



Note: An arrow (→) indicates the direction of a needle inserted percutaneously into the subglottal space towards the interarytenoid area. A cross (x) indicates one point of needle insertion into the interarytenoid by the peroral route.

Fig. 4

There is, however, a possibility of misplacement of the electrodes either in the lateral cricoarytenoid (LCA) if the placement is too deep or in the sternohyoid (SH) if the insertion is too superficial. In order to differentiate the CT from the LCA, the subject is asked to attempt breath-holding or swallowing. These maneuvers should not give EMG activity unless the placement is into the LCA. For discrimination from the SH, the subject is then asked either to open his jaw by resisting the experimenter's hand holding it or to raise his head from the headrest. These attempts will elicit marked activity if the insertion is not deep enough and the electrodes are hooked into the SH.

Thyroarytenoid (VOC). Percutaneous insertion is made with the subject in a supine position attempting sustained phonation. The skin is pierced at a point close to the midline at the level of the cricothyroid space. The needle is then directed cranially and slightly laterally penetrating the cricothyroid membrane to reach the muscle from its inferior surface (Figure 5). This route is slightly different from that reported by Hirano and Ohala (1969), since the needle does not pass through the subglottal space but through the submucous tissues near the anterior commissure.

For verification, the subject attempts to produce low-frequency phonation. The VOC also shows activity during swallowing. Although there is little possibility of contamination with other muscles, the electrodes can pick up the mechanical vibration of the vocal fold if the placement is made too close to the free margin of the fold. In such a case, replacement of the electrode by another insertion is mandatory.

Lateral cricoarytenoid (LCA). The point of insertion is almost the same as for the CT. The needle is then directed laterally and slightly cranially penetrating the cricothyroid membrane at a point anterior to the inferior tuberculum of the thyroid cartilage and deeply enough to reach the LCA. This route is similar to that reported by Hirano and Ohala (1969). (See Figure 5.) Contamination with the VOC can be avoided if the direction of insertion is kept lateral and less cranial so as to stay along the contour of the cricoid ring.

Identification is made by having the subject attempt breath-holding or glottal stop production. These maneuvers, as well as swallowing, give marked activity and serve to discriminate the LCA from the CT as described above.

Strap Muscles of the Neck

Sternohyoid (SH). The contour of the SH can be palpated or even seen through the skin when the subject is asked to raise his head from the headrest in supine position with his head kept extended, unless the subject has a very short, fat neck. The contour is usually clear at the level of the thyroid lamina where the insertion is made. This region is also preferrable to avoid possible contamination with other muscles. As the subject raises his head in a supine position, the needle is inserted lateral to the midline parallel to the alignment of the muscle fibers. This technique is similar to that reported by previous authors (Hirano et al., 1967).

A View of the Larynx with the Right Thyroid Ala Removed

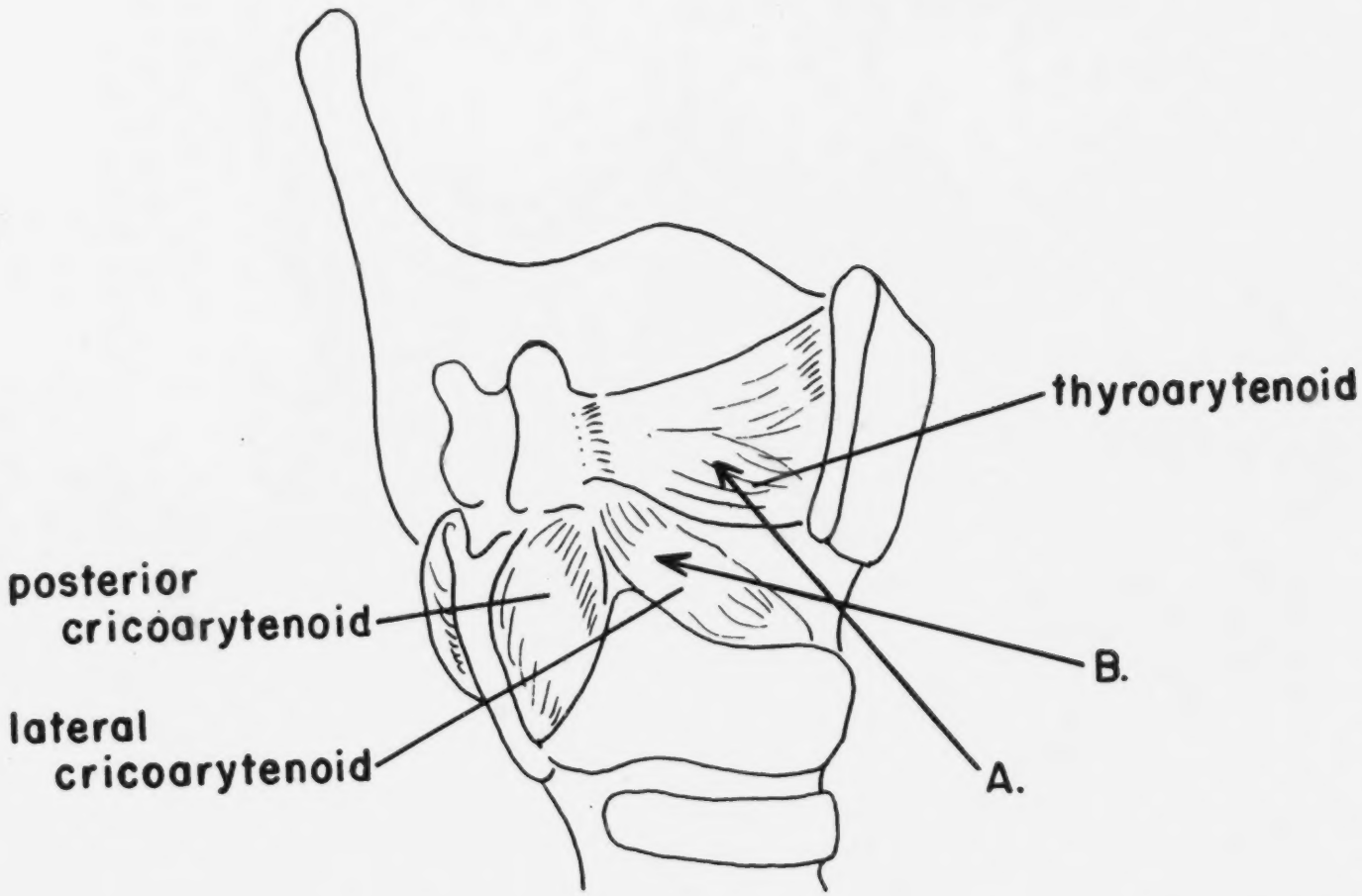


Fig. 5

Note: Arrows indicate the direction of needle insertion into the thyroarytenoid (A) and into the lateral cricoarytenoid (B). (This figure is a modification of Hirano and Ohala, Fig. 6, 1969.)

The exact placement is verified if marked activity is observed when the subject raises his head from the supine position, opens his jaw, or produces very low-frequency phonation.

Sternothyroid (ST). The ST is covered by the SH for almost its entire course in the neck except for the most caudal portion, where it tends to run more medially than the SH, since the ST attaches to the sternum more medially than the SH, as illustrated in standard textbooks of anatomy.

Therefore, our attempt at reaching this muscle is made by inserting the needle at a level 2-3 cm above the suprasternal notch and at the anterior border of the sternomastoid muscle and by directing the needle cranio-laterally. When the subject contracts this muscle by holding his head up from the head-rest in a supine position, we usually feel penetration of its fascia followed by marked EMG activity observed on a monitor oscilloscope. The gross pattern of activity of the ST is not much different from that of the SH so that absolute discrimination from the SH may still be questionable, although it is claimed that the ST appears to be more relevant for pitch lowering than the SH (Simada et al., in press; Hirano, pers. com.).

Thyrohyoid (TH). The TH runs directly on the thyroid lamina and attaches to the linea obliqua, where it is covered by other strap muscles. Insertion of the needle is made at the level of the superior edge of the thyroid lamina, and the needle is pushed caudally and laterally aiming at the linea obliqua until the tip of the needle hits the surface of the cartilage. The cut ends of the electrodes should then be placed in the muscle tissue of the TH, since there is some distance between the hooked ends of the electrodes and the very tip of the beveled end of the needle touching the cartilage.

The EMG activity is observed on a monitor oscilloscope when the subject is asked to attempt quick jaw opening or retraction of the tongue. Again, functional differentiation of the strap muscles is not possible on the basis of present knowledge, so that we must rely on anatomical expectation.

Velopharyngeal Muscles

The peroral approach is always attempted with the subject in a sitting position for insertion into the velopharyngeal muscles. The electrode-bearing needle (usually with an angulated shaft) is held by a pair of alligator forceps.

Levator palatini (LEV). Insertion is made into the levator "dimple" on the soft palate with the subject attempting sustained open vowel phonation. The tip of the needle is directed latero-cranio-posteriorly approximately 10 mm from the surface of the mucosa.

Verification is made by asking the subject to repeat the production of [s]. Marked activity can be observed for this strong oral gesture if the electrodes are placed properly.

Palatoglossus (PG). The PG is the muscular component of the anterior faucal pillar. This muscle is reached by inserting the angulated needle into the anterior pillar either cranio-caudally or in the opposite direction. Since the insertion is made under direct inspection, verification is satisfied if marked activity is shown when the subject swallows.

Palatopharyngeus (PP). In our EMG study, the PP is regarded as the muscular component of the posterior faucal pillar, although there has been some controversy in the past literature on its anatomical description (Bosma and Fletcher, 1962; Fritzell, 1969).

The insertion is made into the posterior faucal pillar under direct inspection. Verification of correct placement is, therefore, satisfied if EMG activity is monitored during swallowing.

Superior constrictor (SC). The tip of an angulated needle is directed cranially to reach the posterior pharyngeal wall lateral to the midline at the estimated level of velopharyngeal closure. The insertion is made under inspection and, therefore, placement is verified if EMG activity is observed for swallowing.

Middle constrictor (MC). Insertion is made using an angulated needle directed caudally into the posterior pharyngeal wall near the level of the tip of the epiglottis. The tongue of the subject is protruded and held for better visualization of the site of insertion.

Practically, precise discrimination of the upper portion of the middle constrictor from the lower fibers of the superior constrictor is difficult, since the constrictor muscles of the pharynx are interlayered at the level of transition from one to the other (Hollinshead, 1966). Therefore, it should be mentioned that what we attempt to examine as the middle constrictor is rather a topographical representation of the pharyngeal constrictor at this particular level. Verification of electrode placement is made in essentially the same way as for the SC.

Suprahyoid and Tongue Muscles

These muscles are reached percutaneously with the subject in a sitting or semi-Fowler position.

Anterior belly of digastric (AD). The contour of the AD is palpable if the subject attempts to open his jaw by resisting the experimenter's hand holding it or if he strongly pushes the tip of his tongue on to the upper alveolar ridge with his mouth slightly open. Insertion is made, with the subject attempting either one of the maneuvers mentioned above, at a point near the anterior attachment of the muscle to the mandibular ridge. The needle is directed obliquely to the surface of the skin aiming at the muscle belly. Verification is made by having the subject open his jaw, a movement which should be accompanied by marked EMG activity. There is little possibility of contamination with other muscles.

Mylohyoid muscle (MH). Insertion into the MH is made near the ridge of the mandible, lateral to the lateral margin of the AD and anterior to the hyoid bone. The experimenter puts his finger onto the floor of the mouth to palpate the tip of the needle perorally. This technique is similar to that reported by Smith and Hirano (1968). (See Figure 6.) Verification is made if the EMG activity is monitored when the subject retracts the tongue backwards or produces [k] repeatedly.

A Frontal Section of the Inferior Portion of the Face Illustrating the Direction of a Needle Inserted into the Mylohyoid (A) and into the Anterior Belly of Digastric (B)

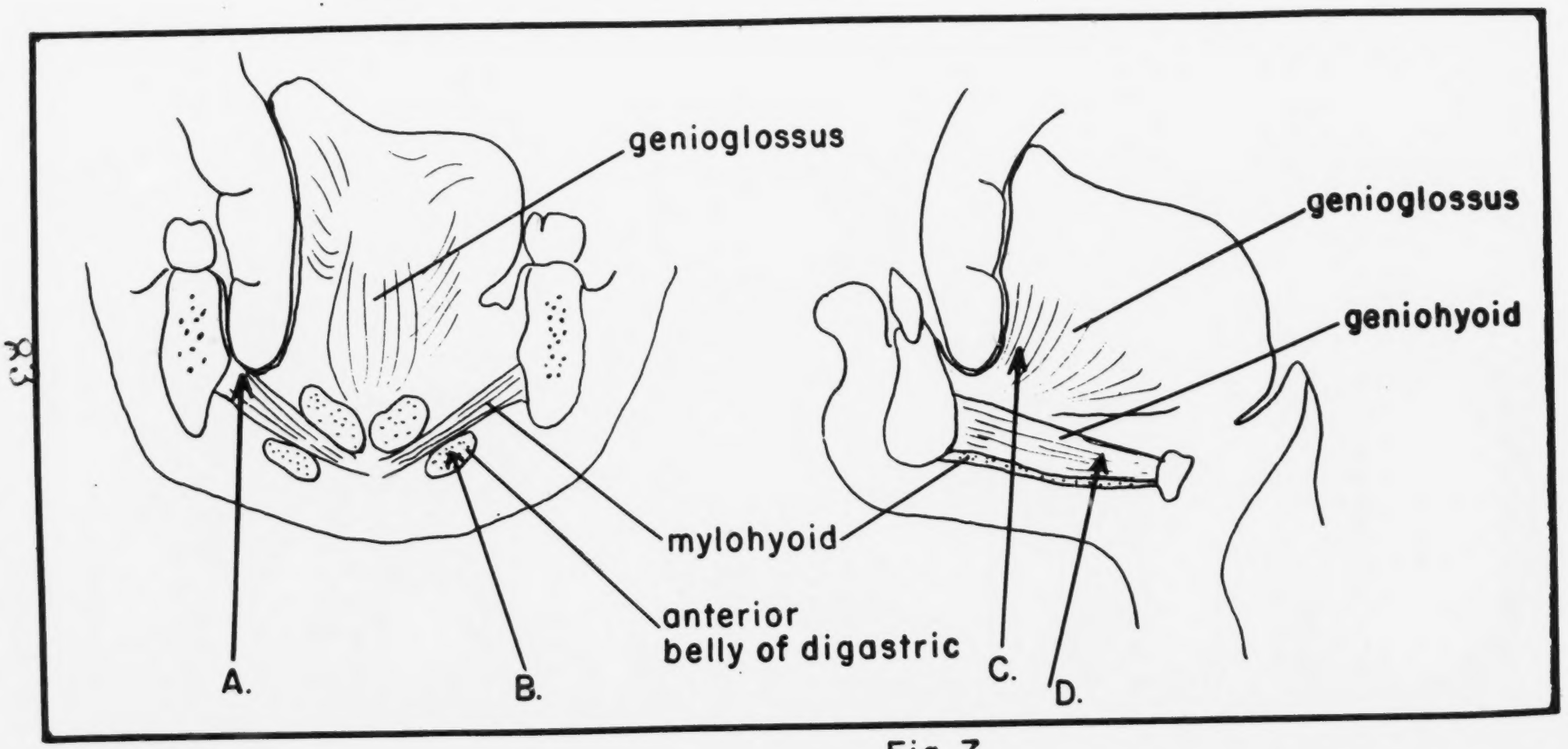


Fig. 6

Note: A fingertip of the experimenter is placed on the floor of the mouth, close to the alveolar ridge, to palpate the tip of the needle. (This figure is a modification of Smith and Hirano, Fig. 1, 1968.)

A Sagittal View of the Inferior Portion of the Face Illustrating the Direction of Needle Inserted into the Genioglossus (C) and into the Geniohyoid (D)

Fig. 7

Note: A finger is palpating to evaluate correct needle placement in the genioglossus. (This figure is a modification of Smith and Hirano, Fig. 2, 1968.)

This muscle is so thin that electrode placement is not always satisfactory in spite of the fact that there is little possibility of contamination with other muscles.

Genioglossus (GG). A percutaneous approach is always taken, although the peroral approach is possible as reported by previous authors (Sauerland and Mitchell, 1970). The needle is inserted perpendicularly to the surface of the skin at the midpoint between the hyoid bone and the mandibular ridge in the paramedian line. The needle is then directed deep enough to be palpated by the experimenter's finger inserted onto the floor of the mouth of the subject. The technique employed in our experiment is the anterior GG placement described by Smith and Hirano (1968). (See Figure 7.)

Exact placement is verified by having the subject protrude his tongue or swallow; vigorous activity should be monitored for these maneuvers. Little possibility of contamination with other muscles is expected if the needle is directed and palpated as described above.

Geniohyoid (GH). The needle is inserted in the paramedian line more caudally than the insertion point for the GG, approximately 10 mm above the level of the hyoid bone. At this level, the MH is almost a tendinous structure covering the inferior surface of the GH. The needle should not be inserted too deeply but should be stopped after the penetration of that tendinous tissue which can usually be felt by the tip of the needle. The estimated depth of insertion is approximately 2-2.5 cm (Figure 7).

Verification of the placement by observing the EMG signal does not seem to be straightforward since there is a conflict of opinions on the function of this muscle (Cunningham and Basmajian, 1969). According to our observation, however, there is some difference in the pattern of the EMG activity of this muscle for swallowing and tongue protrusion of that of the GG, with which the GH is most likely to be confused. Namely, as Cunningham and Basmajian (1969) reported, GH activity follows with some delay in time that of GG activity for the initiation of swallowing, and for simple tongue protrusion, the GH appears to be less active than the GG. Further investigation will be necessary for satisfactory EMG assessment of this particular muscle.

REFERENCES

Basmajian, J.L. and G. Stecko. (1962) A new bipolar electrode for electromyography. J. Appl. Physiol. 17, 849.

Bosma, J.F. and S.G. Fletcher. (1962) The upper pharynx: A review. Part II, Physiology. Ann. Otol. Rhinol. Laryng. 71, 134-157.

Cunningham, D.P. and J.V. Basmajian. (1969) Electromyography of genioglossus and geniohyoid muscles during deglutition. Anat. Rec. 165, 401-409.

Fritzell, B. (1969) The velopharyngeal muscles in speech. An electromyographic and cineradiographic study. Acta Otolaryng., Suppl. 250, 1-81.

Gaskil, J.R. and D.D. Gillies. (1966). Local anesthesia for peroral endoscopy. Arch. Otolaryng. 84, 654-657.

Harris, K. (1970) Physiological measures of speech movements: EMG and fiberoptic studies. Proceedings of the Workshop on Speech and the Dentofacial Complex. ASHA Reports No. 5, 271-282.

Hirano, M., Y. Koike and H. van Leden. (1967) The sternohyoid muscle during phonation: Electromyographic studies. Acta Otlaryng. 64, 500-507.

Hirano, M. and J. Ohala. (1969) Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. JSHR 12, 362-373.
Hollinshead, V.H. (1966) Anatomy for Surgeons. (Vol. I: Head and Neck.) New York: Harper and Row.
Sauerland, E.K., and S.P. Mitchell. (1970) Electromyographic activity of the human genioglossus muscles in response to respiration and to positional changes of the head. Bulletin of the Los Angeles Neurol. Soc. 35, 69-73.
Shipp, T. (1968) A technique for examination of laryngeal muscles during phonation. Proc. 1st Int. Cong. of Electromyographic Kinesiology, Montreal, Canada, August 1968. 21-26.
Shipp, T., W.W. Deatsch and K. Robertson. (1968) A technique for electromyographic assessment of deep neck muscle activity. Laryngoscope 78, 418-432.
Simada, Z., Hirose, H., M. Sawashima and O. Fujimura. (In press) An EMG study of Japanese accent patterns. Proc. for 7th International Congress on Acoustics, Budapest, 1971.
Smith, T. and M. Hirano. (1968) Experimental investigation of the muscular control of the tongue in speech. Working Papers in Phonetics UCLA. 10, 145-155.
Zemlin, W.R. (1969) The effect of topical anesthesia on internal laryngeal behavior. Acta Otolaryng. 68, 196-176.

Action of the Extrinsic Musculature in the Control of Tongue Position: Preliminary Report

Katherine S. Harris*
Haskins Laboratories, New Haven

The position of the tongue in the mouth is controlled, in part, by a group of muscles which connect it to the mandible and the hyoid bone. In addition, since the tongue itself rests on the hyoid, its position is influenced by forces acting on the hyoid. While the possible functions of the muscles can be inferred from their origins and insertions, as described in the usual anatomical texts, the tongue's position in running speech depends crucially on muscle interactions, which must be directly observed.

The purpose of the present study was twofold: first, to supplement existing normative data on speech function and, second, to continue work begun, particularly by MacNeilage and deClerk (1969) and Smith and Hirano (1968), on the difficult problem of understanding positional variants of the phoneme. Some extremely preliminary results will be reported here.

METHOD

Electrodes were inserted into the genioglossus and various infra- and suprahyoid muscles, by the techniques described by Hirose (1971). Two subjects were used; most of the data reported here are from the second run of one subject. The subjects read random lists of the form /ə CVC/. The first consonant was /p/, /t/, or /k/; the second consonant was /p/, /t/, or /k/; and the vowel was /i/, /ɑ/, or /u/.

Output data processing is described by Port (1971). For averaging, the utterances were lined up at the end of voicing for the initial unstressed vowel. In the following figures, line-up is indicated by "0" on the abscissa, while onset and termination of voicing for the stressed vowel are indicated by arrows.

RESULTS

Genioglossus

The output of the genioglossus is shown for the syllables /pip/, /pɑp/, and /pup/ in Figure 1. As one might expect from traditional descriptions of the function of the muscle, activity is greatest for /i/, less for /u/, and least for /ɑ/. No distinct peaks were associated with any initial or terminal consonants, although there was some modification of the rise and fall contours of the vowel; the peak heights of the vowels were not influenced by initial and terminal consonants. Our results on this point seem to be roughly comparable with those of Smith and Hirano (1968), although it is difficult to be sure without an opportunity to make more detailed comparisons.

*Also The Graduate Center of the City University of New York

Genioglossus

/pip/ ·····
/pap/ ——
/pup/ - - - - -

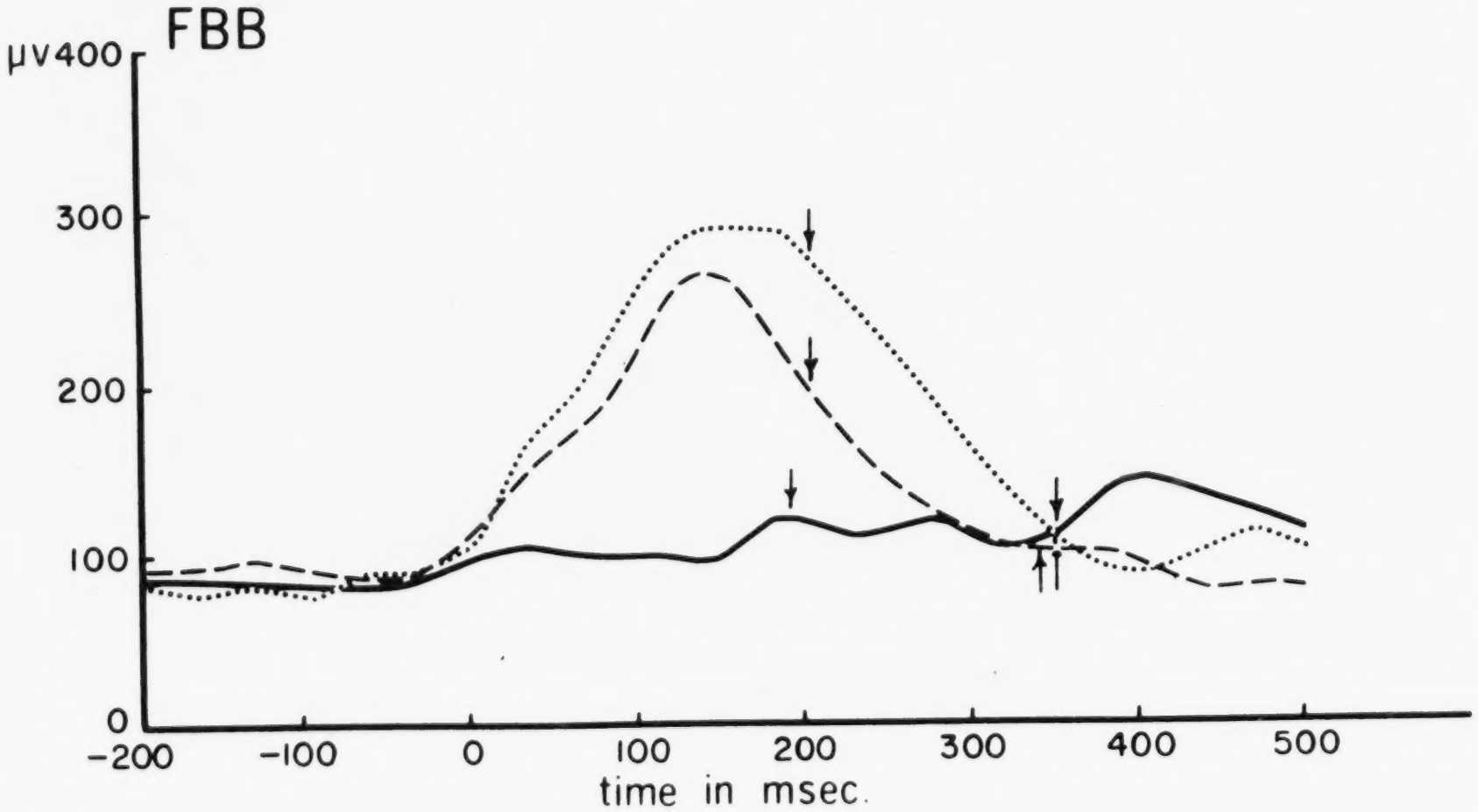


Fig. 1

Anterior Belly of the Digastric

The action of the anterior belly of the digastric is quite clearly to open the jaw. As shown in Figure 2, there is essentially no action for /i/ and /u/ and a large peak for /ɑ/. Peak sizes for the vowel do not seem to be affected by preceding or following consonants. A similar result has been reported by Hirose et al. (1968, 1969).

Strap Muscles

Data from the same three CVC syllables for the sternohyoid, sternothyroid, and thyohyoid muscles are shown in Figures 3, 4, and 5, respectively. All three appear to be correlated with jaw opening, as is the action of the anterior belly. This result has been previously reported by Ohala and Hirose (1970) and by Gårding et al. (1970). There is some tendency for the peak to be somewhat larger for /i/ than for /u/. However, the data from the thyrohyoid and sternothyroid muscles are most unsatisfactory from the point of view of recording level.

Mylohyoid

The activity of the mylohyoid for the same three utterances is shown in Figure 6. Here /ɑ/ and /u/ show similar patterns, while /i/ is considerably higher. This pattern seems in general agreement with the presumed function of the mylohyoid in raising the floor of the mouth, although the difference between /i/ and /u/ is not explicable on this simple basis. Smith and Hirano (1968) report no activity for any vowel in these environments, which is somewhat puzzling.

The mylohyoid is unlike the other muscles described here in that much more substantial peaks are seen for the consonant [k] than for any vowel and that [t] is also quite large, whether in initial or terminal position. Figure 6 shows an example of this sort. Detailed comparisons of peak sizes can be made by examining Table 1.

As one can see, there are modifications of peak size in the terminal consonant depending on the preceding vowel and modification of the vowel depending on the preceding consonant. In addition, there are modifications of the size of the initial consonant peak depending on the following vowel. Similar, although not identical, results are reported by Smith and Hirano (1968).

COMMENTS

The types of interaction reported here have been previously discussed by MacNeilage (1970), MacNeilage and deClerk (1969), and Smith and Hirano (1968). The modification of the terminal peak by the preceding vowel and the modification of the vowel by the preceding consonant are MacNeilage and deClerk's left-to-right effects and are quite common in EMG studies, as they point out. They do not, however, necessarily represent a modification of target position in movement terms but may merely reflect the fact that muscle contractions will be larger if more movement of the articulators is required. Even if, for example, an X-ray study of the tongue showed the same position for [i] after [k] as after [p], we would expect to find left-to-right effects at the EMG level.

The modification in size of the initial [k] and [t] peaks is a right-to-left effect, sometimes described as anticipatory coarticulation. Anticipatory

Anterior belly

/əpip/ .....
/əpɑp/ - - - - -
/əpup/ ——

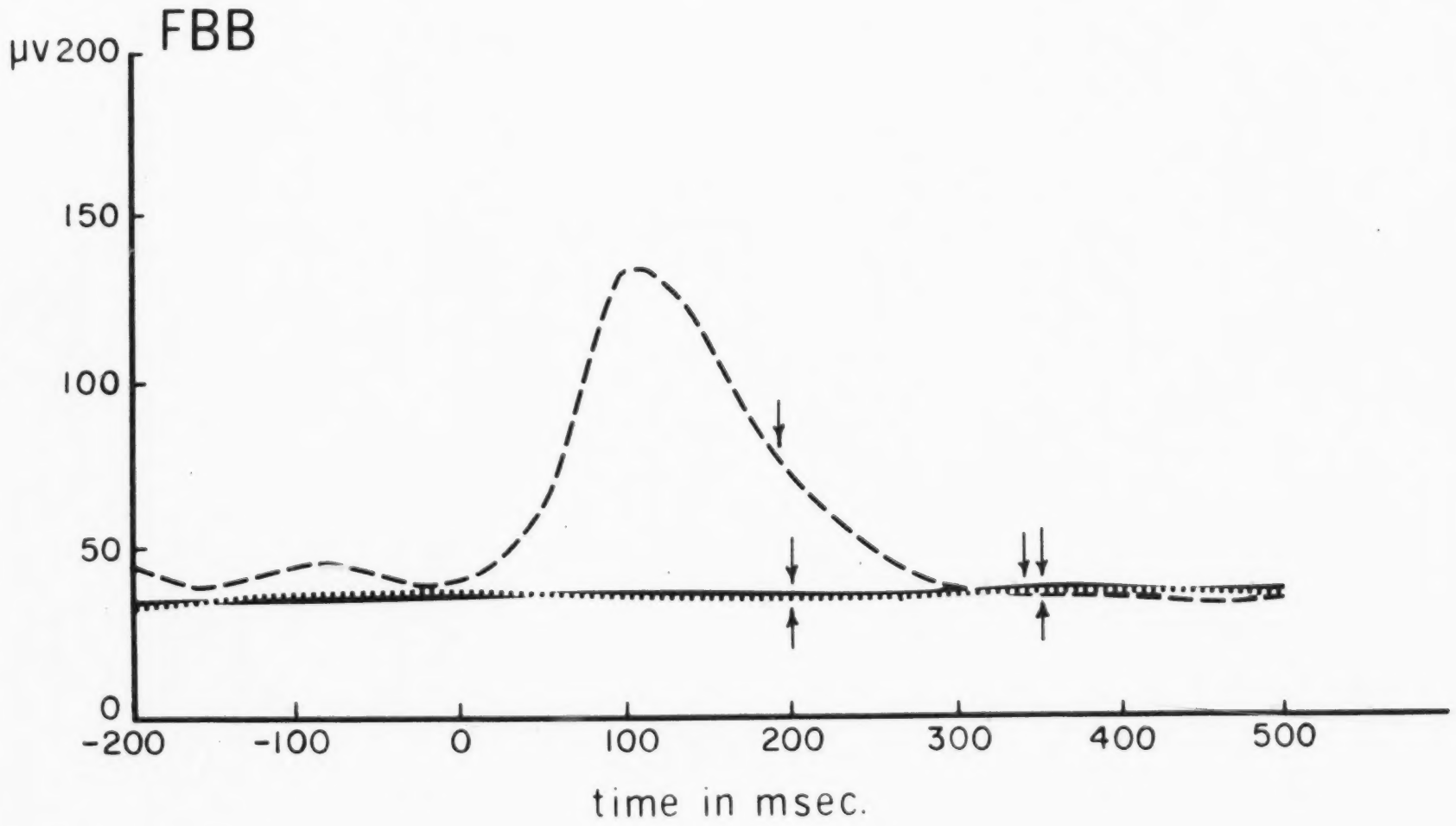


Fig. 2

Sternohyoid

/əpip/ .....
/əpɑp/ - - - - -
/əpup/ ——

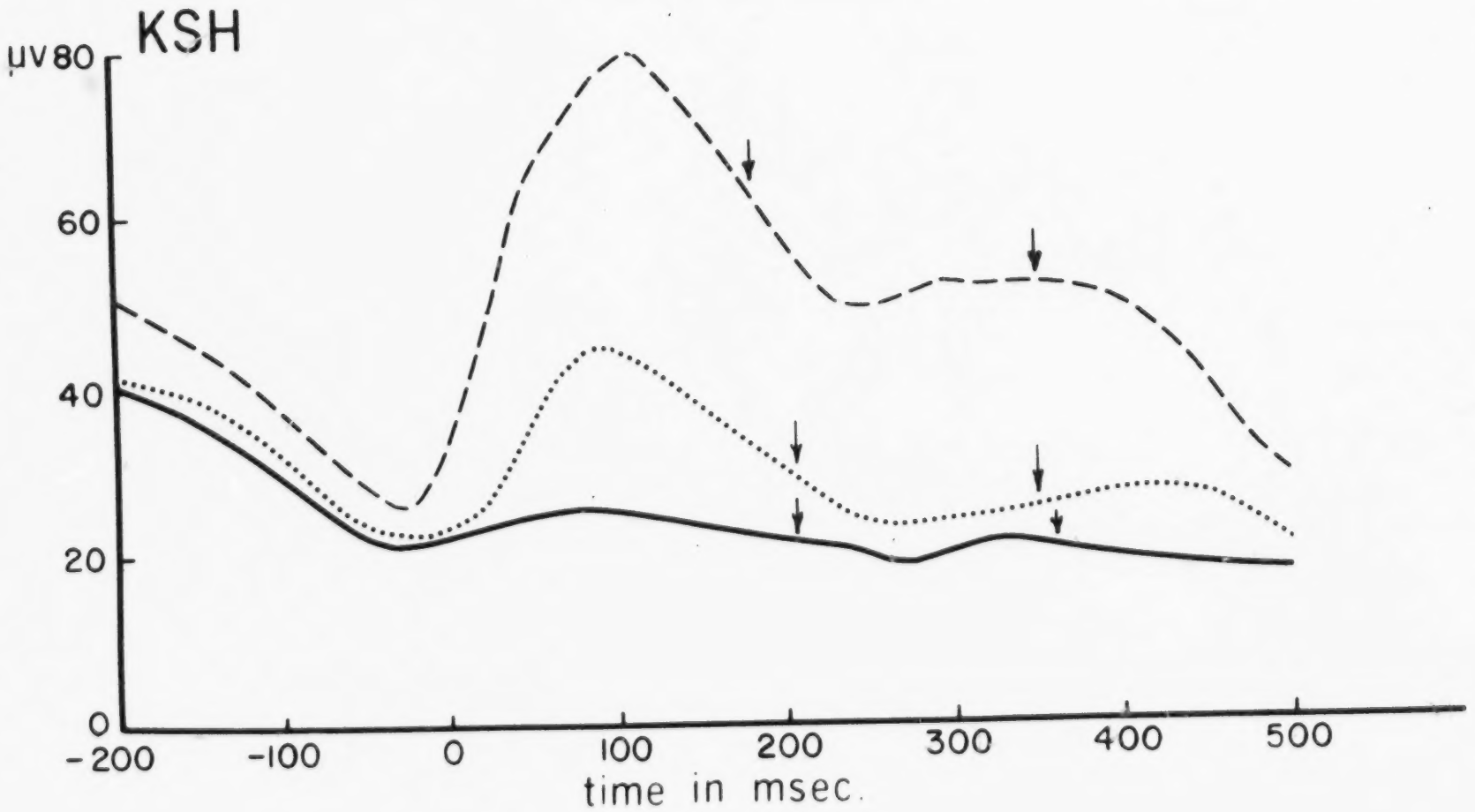


Fig. 3

Sternothyroid

/əpip/ .....
/əpɑp/ - - - - -
/əpup/ ——

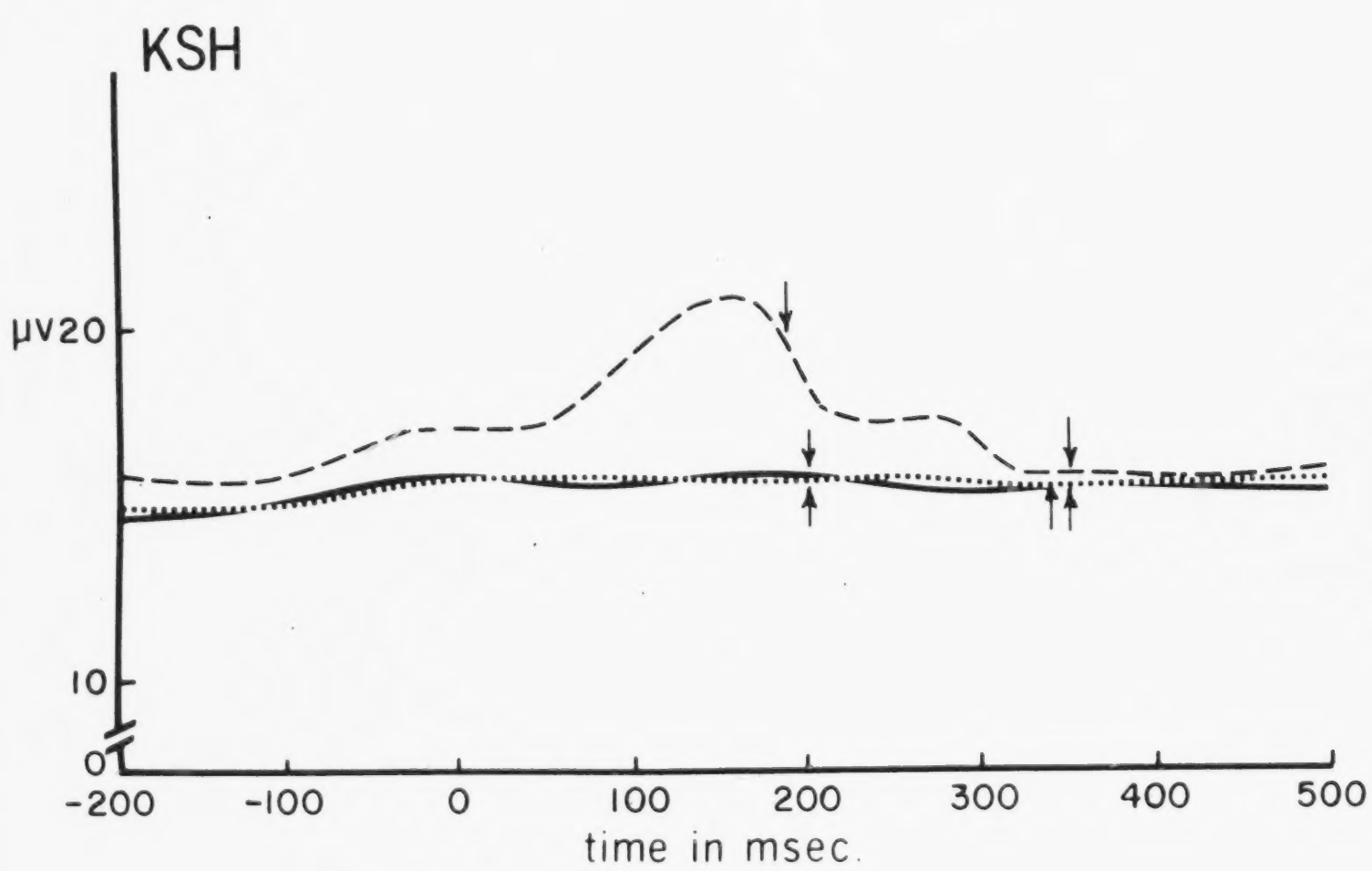


Fig. 4

Thyrohyoid

/əpip/ .....
/əpɑp/ - - - - -
/əpup/ ———

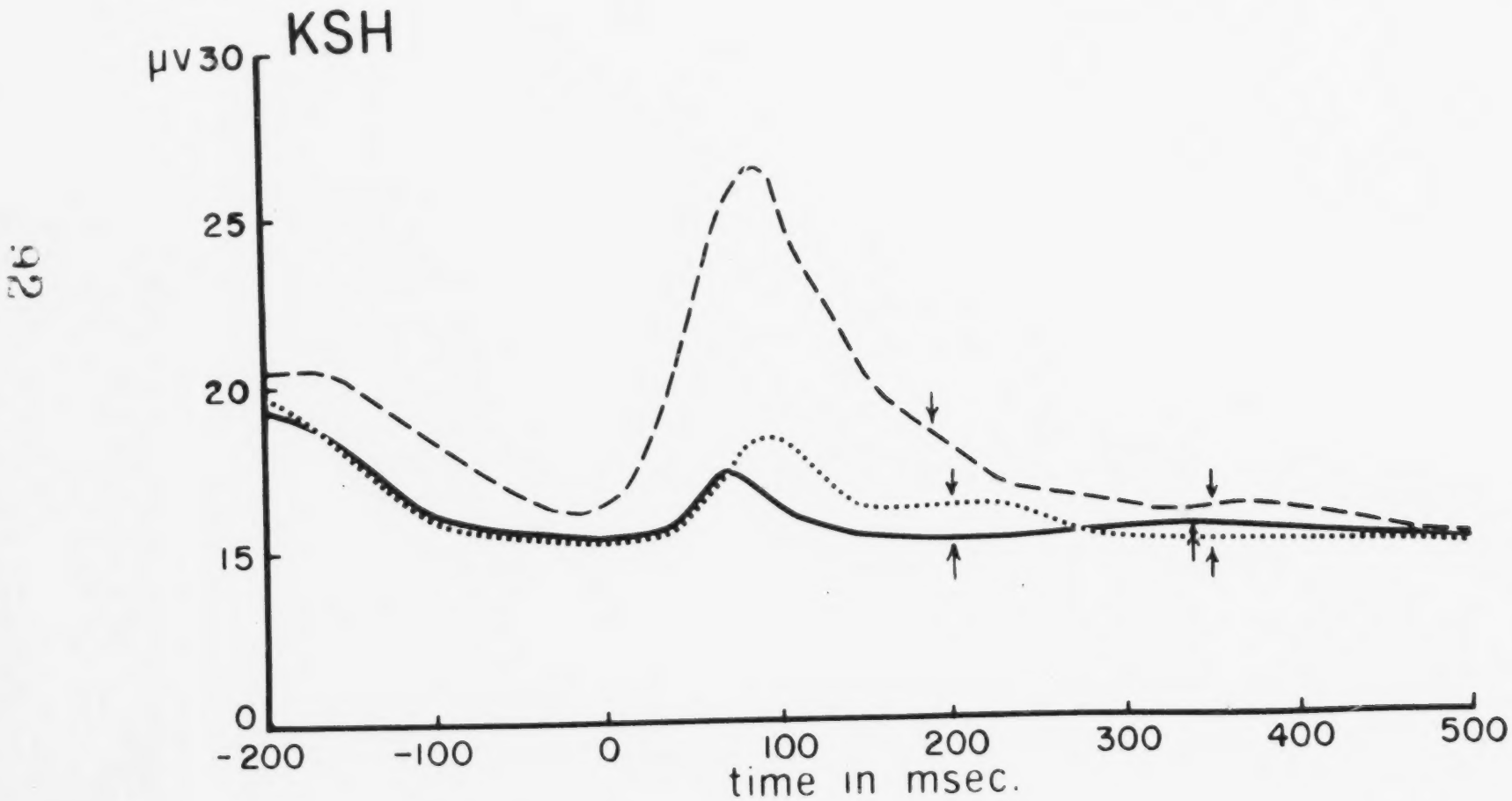


Fig. 5

Mylohyoid

/pik/ .....
/pak/ ——
/puk/ - - - - -

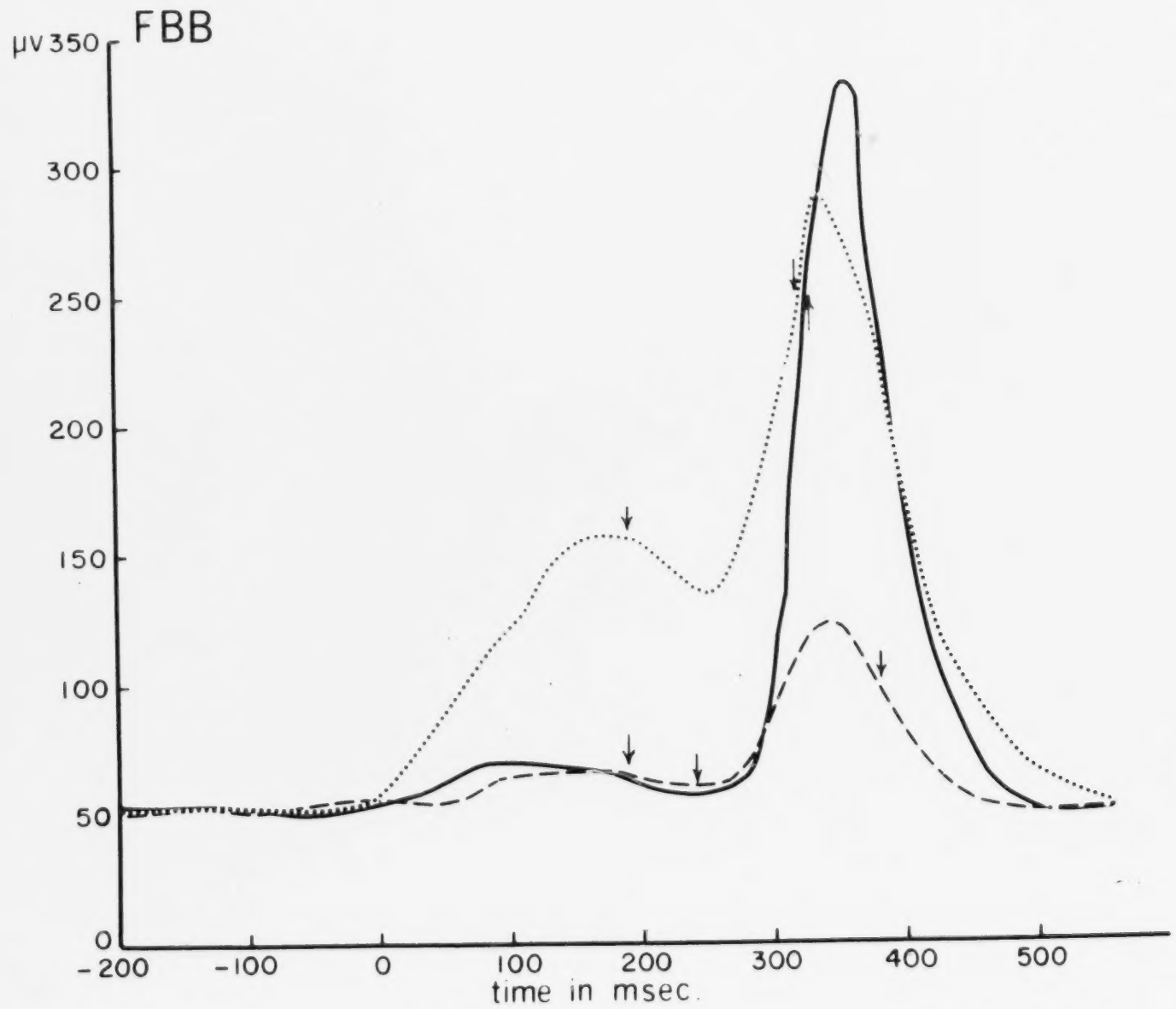


Fig. 6

Table I

	P_1	P_2	P_3		P_1	P_2	P_3		P_1	P_2	P_3
pip	50	150	50	tip	100	150	60	kip	360	135	80
pit	60	170	55	tit	105	160	120	kit	320	130	115
pik	50	155	290	tik	130	150	210	kik	310	145	270
pap	50	80	55	tap	195	80	60	kap	435	60	75
pat	60	85	55	tat	150	75	65	kat	490	65	80
pak	55	70	330	tak	175	75	325	kak	480	65	320
pup	50	75	50	tup	130	90	65	kup	390	55	55
put	55	80	75	tut	115	90	120	kut	385	60	155
puk	55	70	125	tuk	115	75	175	kuk	435	55	190

coarticulation has been studied before both at the EMG level and at the movement level (see Amerman et al., 1970, for example). However, two rather different kinds of phenomena are described this way. The common example given is the rounding of the lips during [t] closure, when the following vowel is [u]. This may represent simply a change in timing of the activity associated with the vowel and does not necessarily indicate a change in the muscular organization of vowel formation.

The example here is quite different; we might hypothesize that the [k] peak is made by the combined action of the genioglossus and the mylohyoid (and probably other muscles) when the genioglossus is active for the vowel but by the mylohyoid when the genioglossus is not active for the vowel and when, in addition, the jaw is opening for the vowel. Although mere timing changes in muscle action could explain the lip-rounding example, they cannot explain this type of reorganization. Indeed, any explanation must depend on some complex type of preprogramming, since reorganization depends on events from which feedback is not yet available. Obviously, these phenomena will require detailed study in the future.

REFERENCES

Amerman, J. D., R. Daniloff, and K. L. Moll. (1970) Lip and jaw coarticulation for the phoneme /ɹ/. JSHR 13, 147-161.

Gårding, E., O. Fijimura, and H. Hirose. (1970) Laryngeal control of Swedish word tones: A preliminary report on an EMG study. Ann. Bull. Inst. Logo. and Phoniatr. 4, 45-54.

Hirose, Hajime. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Hirose, H., S. Kiritani, and S. Sibata. (1968) An electromyographic study of articulatory movements. Ann. Bull. Res. Inst. Logo. and Phoniatr. 2, 21-27.

Hirose, H., Z. Simada, and S. Kiritani. (1969) An electromyographic study of articulatory movements. Ann. Bull. Res. Inst. Logo. and Phoniatr. 3, 42-55.

MacNeilage, Peter F. (1970) Motor control of serial ordering speech. Psychol. Rev. 77, 182-195.

MacNeilage, P.F. and J.L. deClerk. (1969) On the motor control of coarticulation in CVC monosyllables. J. Acoust. Soc. Amer. 45, 1217-1233.

Ohala, J., and Hirose, H. (1970) The function of the sternohyoid muscle in speech. Ann. Bull. Inst. Logo. and Phoniatr. 4, 41-44.

Port, D. K. (1971) The EMG data system. (See this Status Report.)

Smith, T. and M. Hirano. (1968) Experimental investigation of the muscular control of the tongue in speech. Working Papers in Phonetics 10, 145-155.

Electromyography of the Intrinsic Laryngeal Muscles During Phonation

Thomas Gay,[+] Hajime Hirose,[++] Marshall Strome,[+++] and Masayuki Sawashima[++++]

Electromyographic studies of the laryngeal muscles during phonation have been widely reported in the literature, with the classic experiments of Faaborg-Andersen (1957, 1965), in particular, providing a basis for describing the laryngeal control of phonation. Nonetheless, a number of questions about the control of fundamental frequency and intensity within and across vocal registers and the reliability of EMG measures, in general, have remained unanswered. This was due, largely, to the technical problems inherent in using concentric needle electrodes and the difficulty in extracting subtle changes in muscle activity patterns from raw EMG data. However, recent advances in both EMG recording and processing techniques have provided the necessary capability for answering these questions. On the one hand, hooked-wire electrode insertion techniques (Hirano and Ohala, 1969) have made possible the simultaneous recording of the intrinsic laryngeal musculature with a minimum of equipment interference and subject discomfort. On the other hand, the use of a digital computer to average the integrated curves of a number of tokens of a given vocal maneuver (Cooper, 1965; Gay and Harris, in press) has provided a convenient and accurate means of displaying the average strength of contraction of a given muscle or muscle group.

The primary purpose of this experiment was to describe, in detail, the actions of the intrinsic laryngeal muscles during various vocal frequency- and intensity-changing maneuvers. In addition, the conditions of the experiment were designed to simulate those of an earlier study (Sawashima et al., 1969) in order to obtain data on the reliability of repeated EMG measurements.

PROCEDURES

Subjects were five adults, four male and one female, all native speakers of American English. The female subject was a trained singer.

For each subject, an attempt was made to record from the five intrinsic muscles simultaneously. However, this goal was reached only for two of the five subjects. Unsatisfactory recordings were obtained for the vocalis[1]

[+]Haskins Laboratories, New Haven, and Department of Oral Biology, University of Connecticut Health Center, Farmington.

[++]Haskins Laboratories, New Haven, and Faculty of Medicine, University of Tokyo.

[+++]Department of Otolaryngology, School of Medicine, Harvard University, Cambridge.

[++++]Research Institute of Logopedics and Phoniatrics, University of Tokyo.

[1]By reason of both past experience and the verification techniques employed, we are confident that we isolated the vocalis muscle. However, since the insertion was not viewed directly, we cannot be virtually certain that the electrode field did not include any potentials from the "external" thyroarytenoid.

muscle of one subject, for the interarytenoid muscle of another, and for the posterior cricoarytenoid and cricothyroid muscles of the third.

The EMG data were collected by following our usual procedures of hooked-wire electrodes, after the type described by Basmajian and Stecko (1962), and computer processing (Hirose, 1971; Port, 1971).

The acoustic measurements of fundamental frequency and relative intensity were made from oscillographic records obtained from a Honeywell Visicorder optical oscillograph.

Electromyographic data were collected for three different conditions of phonation:

1) Frequency Control: Chest Register - a stepwise change in fundamental frequency (as an arpeggio, "do-mi-sol-do-sol-mi-do") for phonation of a sustained /a/ at both moderate and loud intensity levels.

2) Frequency Control: Falsetto - sustained phonation of /a/ at high pitch-chest register, low pitch-falsetto, high pitch-falsetto.

3) Intensity Control - sustained phonations of /a/ for combinations of three pitch conditions (low pitch-chest, high pitch-chest, falsetto) and three intensity conditions (low, moderate, high).

4) Vocal Attack - sustained phonation of /a/ with three different attacks: breathy, simultaneous, glottal. (Data not presented here.)

All utterances were repeated successively between ten and twenty times. For each trial of frequency control, subjects were instructed to keep constant intensity regardless of the change in frequency of voice. The subjects were allowed ample practice and, in addition, were able to monitor intensity levels by means of a db meter. In the intensity control experiment, the subjects were asked to phonate at three different intensity levels for each fundamental frequency level, maintaining a constant fundamental frequency for each intensity level. Where necessary, the subjects used earphones to match their fundamental frequencies to the output of a sine wave oscillator.

RESULTS AND DISCUSSION

Frequency Control: Chest Register

In general, the data of this series show that increases in fundamental frequency are accompanied by progressive increases in the activity of the tensor muscles of the larynx. This is clearly illustrated in Figure 1, which summarizes the low-intensity arpeggio data for a single male subject. Since the averaged EMG curves remained at a relatively steady level throughout the duration of each arpeggio step (except for some overshoot at the onset of phonation), each step is shown by a single data point, which represents the graphic average (straight line fit) of the curve between 300 and 500 msec after the onset of phonation. Although an increase in fundamental frequency produces an increase in the activity of all intrinsic muscles, the greatest increase is for the cricothyroid and vocalis muscles. Note also the activity of the posterior cricoarytenoid, which increases markedly at the highest pitch level. Apparently, the posterior cricoarytenoid, as an antagonist to the cricothyroid, can also act as a tensor of the vocal folds. Figure 2 shows the same data for the high-intensity arpeggio. Here, the same activity pattern is evident but

Low-Intensity Arpeggios

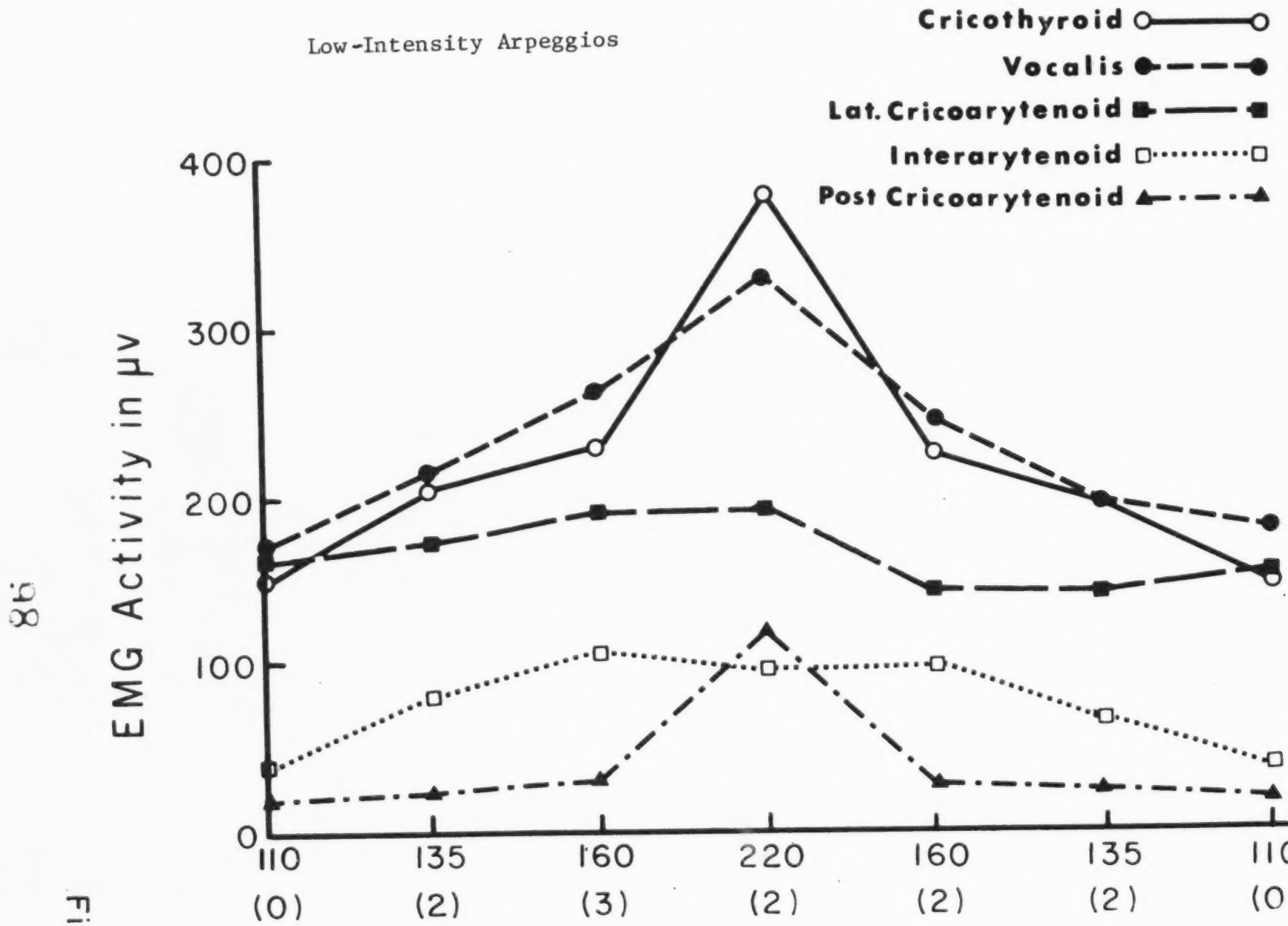


Fig. 1

Note: Points along the curves represent averages of EMG data for the fundamental frequencies noted along the abscissa for Subject LJR. Intensity levels (in db) relative to the first arpeggio step (=0) are shown beneath the frequency values.

High-Intensity Arpeggios

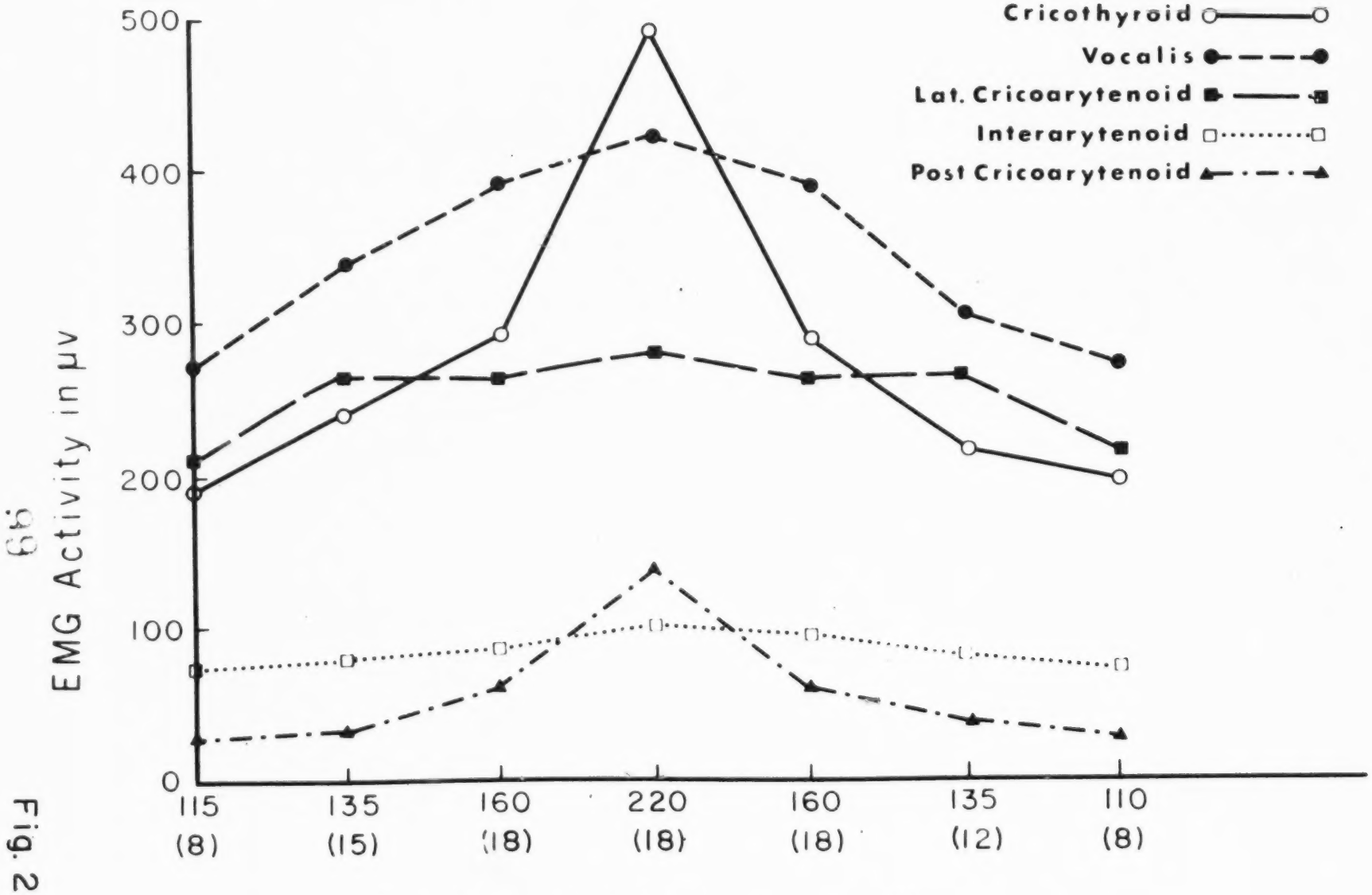


Fig. 2

Note: Points along the curves represent averages of EMG data for the fundamental frequencies noted along the abscissa for Subject LJR. Intensity levels (in db) relative to the first arpeggio step (=0) are shown beneath the frequency values.

with higher levels for both the tensor muscles and the interarytenoid muscle. Posterior cricoarytenoid activity is also apparent, following the curve of the cricothyroid.

With respect to the cricothyroid, vocalis, and posterior cricoarytenoid muscles, the data obtained from the other subjects showed quite similar activity patterns, with progressive increases of activity accompanying stepwise increases in fundamental frequency and a general heightening of overall tensor activity for the higher intensity series. However, some variability was found for the adductor muscles. The increase in activity for the interarytenoid muscle at high intensity was peculiar to this subject. Other subjects also showed individual patterns of adductor muscle activity. One subject, for example, showed a marked increase in lateral cricoarytenoid activity at only the highest arpeggio step for both intensity conditions. Generally though, the higher frequency steps were characterized by only slight increases in adductor activity.

It is generally agreed that the cricothyroid and vocalis muscles are primarily responsible for the control of fundamental frequency. The data of this experiment show, further, that the actions of the two muscles vary systematically with both upward and downward changes in fundamental frequency. It has also been suggested (Sawashima et al., 1969) that the functions of these two muscles in regulating fundamental frequency differ in that the activity for the cricothyroid muscle varies more linearly with changes in frequency. The data obtained here show, rather, similar changes in activity patterns for both muscles. In a strict sense though, neither seems to bear a linear relationship to fundamental frequency.

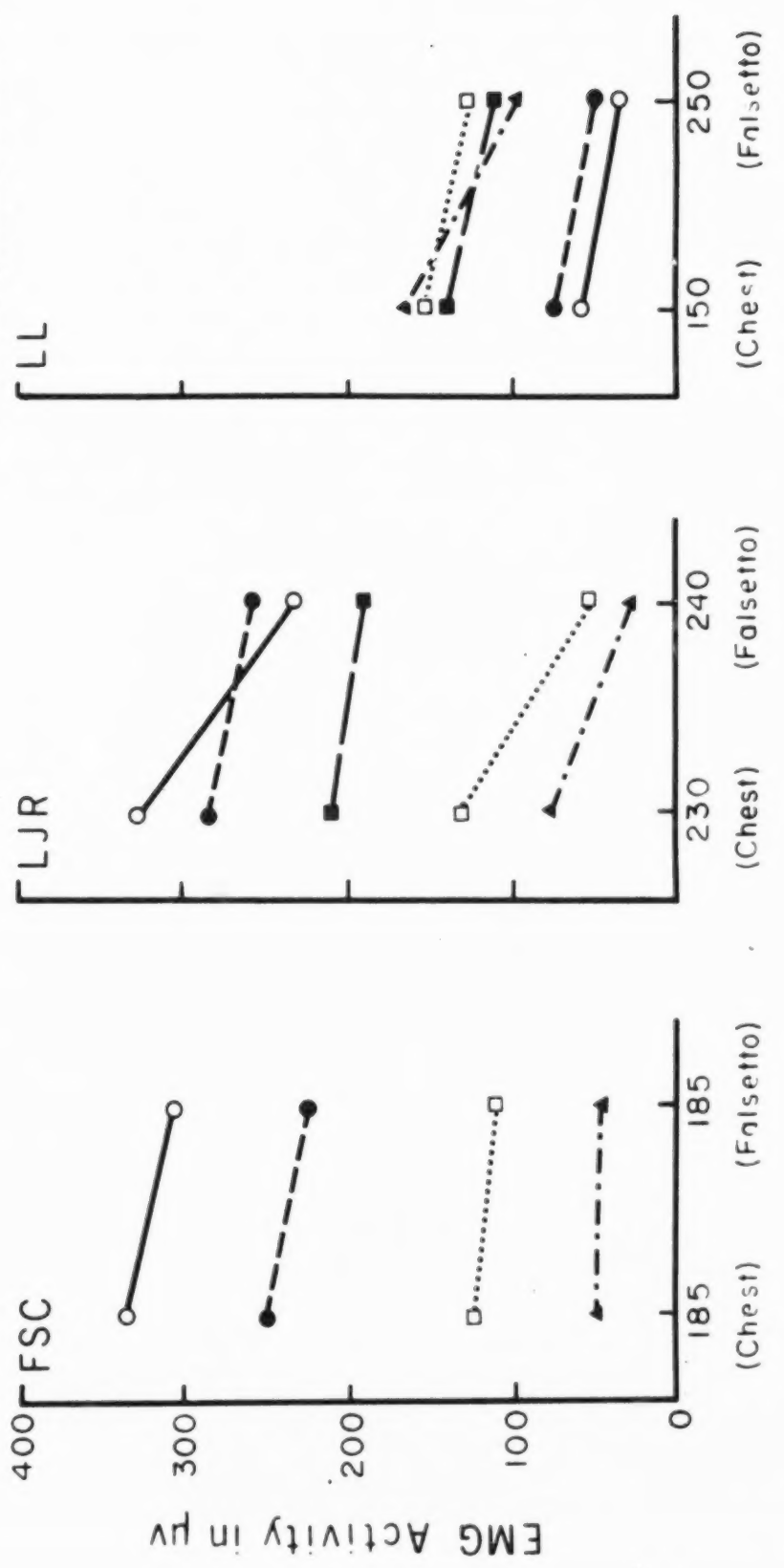
The posterior cricoarytenoid finding is an interesting one and one which is in disagreement with the data of Faaborg-Andersen (1957), which showed relaxation of the posterior cricoarytenoid with increases in fundamental frequency. The contribution of the adductor muscles to changes in fundamental frequency is also less than straightforward. Hirano, Ohala, and Vennard (1969) suggest that the lateral cricoarytenoid participates in the regulation of fundamental frequency. The data of this experiment show that, indeed, the lateral cricoarytenoid sometimes does show increased activity with increases in pitch, but its actions, when evident, seem less consistent than those of the tensors. The interarytenoid reveals the same variability, depending on the particular subject.

Briefly summarizing then, the dominant muscle forces in regulating fundamental frequency in chest register are those of the cricothyroid and vocalis, with some antagonistic action of the posterior cricoarytenoid, especially at the higher frequency levels. Adductor muscle action probably plays a secondary role with specific contributions varying with the individual.

Frequency Control: Falsetto

Previous experiments (Faaborg-Andersen, 1957, 1965; Hirano et al., 1969; Sawashima et al., 1969) have shown that vocalis muscle activity (and often cricothyroid muscle activity) decreases with a shift in register from chest voice to falsetto. The data shown in Figure 3 confirm this and indicate, moreover, that a shift from high chest voice to low falsetto is reflected by a generalized relaxation of all the laryngeal muscles. However, increases in pitch

Sustained Phonation - Different Registers



Note: Points represent averages of EMG data for fundamental frequencies in the registers noted and shown along the abscissa, for three subjects.

Fig. 3

within falsetto were accompanied by greater overall muscle activity. In the case of the trained singer, the muscle activity pattern for an arpeggio sung entirely in falsetto mirrored the pattern for chest voice, but with a lower corresponding level of muscle activity, i.e., the average EMG level for the first step in falsetto (260 Hz) was lower than that for the highest step (also 260 Hz) in the chest voice arpeggio.

Intensity Control

Generally speaking, the regulation of vocal intensity can be accounted for by changes in glottal resistance (laryngeal tension), by subglottal air pressure, or by both. As with previous EMG studies of intensity control, the data of this study provide direct information on only the laryngeal tension aspect; subglottal pressure contribution can be made only by inference.

The results of earlier EMG studies of the larynx are somewhat contradictory regarding the mechanism of intensity control. Both Faaborg-Andersen (1957) and Sawashima et al. (1958) report no significant change in the activity of the vocalis or cricothyroid muscles with changes in intensity, while Hirano et al. (1969) suggest active participation of the vocalis and lateral cricoarytenoid in regulating intensity in chest register, with a reduction of activity in falsetto.

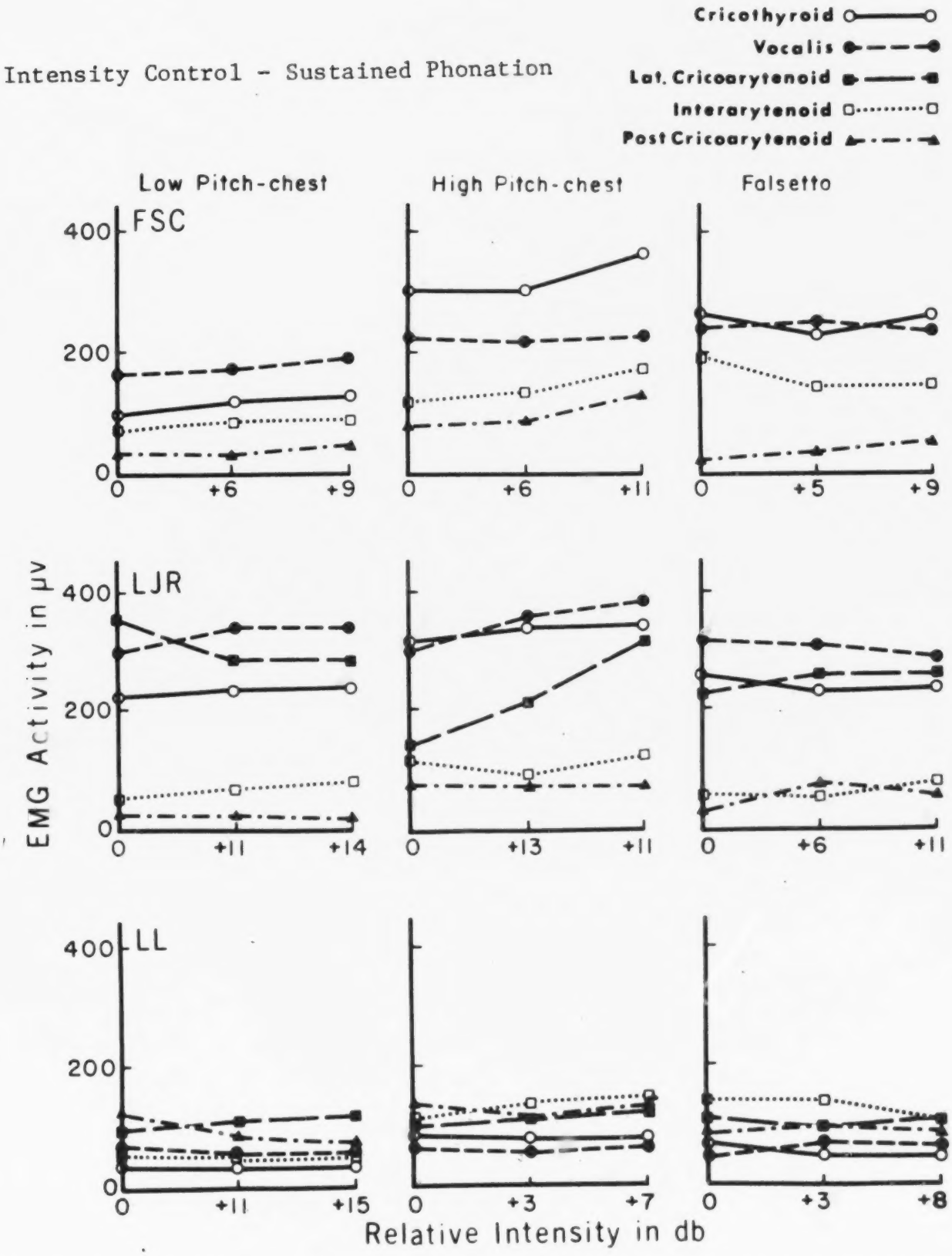
In this series, EMG data were obtained for combinations of three pitch conditions (low-chest, high-chest, falsetto) and three intensity conditions (low, moderate, high). Figure 4 summarizes the data for three subjects. Again, each data point represents the averaged muscle activity between 300 and 500 msec after the onset of phonation.

The top row of Figure 4 summarizes the intensity data for Subject FSC. At low pitch-chest, there are only very slight increases in muscle activity across changes in intensity. At high pitch-chest, activity increases are sharper for the cricothyroid, lateral cricoarytenoid, and posterior cricoarytenoid, but vocalis activity levels off. There is a general leveling off or reduction for all muscles in falsetto. The curves for LJR show less general increase, except for vocalis and interarytenoid activity in high pitch-chest. The curves for LL, on the other hand, are relatively flat for all sets, with even some reduction of activity at high-intensity falsetto.

Except in three instances, muscle activity levels remained relatively steady or increased only slightly across changes in vocal intensity. Levels for falsetto were especially steady. Larger increases are more evident among sets, that is, as a function of fundamental frequency change. Also, given even the slight increases related to intensity, it would seem unlikely that the small changes in activity observed could be responsible for the large increases in intensity levels produced.

Another finding is worth mentioning. In a previous study, Hirano et al. (1969) found that cricothyroid activity decreased as vocal intensity increased. They suggested that this is a compensatory mechanism for regulating fundamental frequency under conditions of increased laryngeal tension (high intensity). This pattern of muscle activity was not evident for any of the present subjects. Generally, cricothyroid activity either leveled off or increased slightly across increases in intensity. However, since Hirano, Ohala, and

Intensity Control - Sustained Phonation



Note: Points in each graph represent averages of EMG data for three intensity levels. Data are shown for three pitch levels (chest register-low pitch; chest register-high pitch; falsetto) for each of three subjects. Intensity levels (in db) are relative to the lowest intensity level produced for each frequency (=0) and are shown along the abscissa. Fundamental frequencies are:

	Register		
Subject	Chest-Low	Chest-High	Falsetto
FSC	105 Hz	190 Hz	200 Hz
LJR	130 Hz	180 Hz	320 Hz
LL	95 Hz	190 Hz	290 Hz

Fig. 4

Vennard's subjects produced swelltones while the present subjects phonated steady-state vowels, both results are probably equally tenable, if the contextual differences are taken into account.

Reliability of Repeated Measurements

As was mentioned at the onset, the conditions of this experiment were designed to simulate those of an earlier one on the tensors of the larynx by Sawashima et al. (1969). Two subjects in that experiment were also subjects in the present one.

The arpeggio data for both subjects were quite consistent across the two experiments. Although actual levels differed, activity changes were always systematic. This was further confirmed when a second opportunity arose during the course of this experiment to obtain another set of arpeggio data for one of the subjects (LJR). Again, analysis showed systematic changes in tensor muscle activity for stepwise changes in fundamental frequency along with increased activity of the posterior cricoarytenoid at the highest pitch levels.

Tensor muscle relaxation accompanying a shift to falsetto was also consistent for the two studies. Unfortunately, since much of the intensity and voice onset data were fragmentary, other meaningful comparisons could not be made. One final comparison, though, was possible. In the present experiment, Subject TG was one of two who showed a large peak in vocalis activity for glottal attack. This was the same pattern evident in the first experiment.

These similarities are interesting, especially in light of the fact that different electrodes were used for the first experiment (concentric needle as opposed to hooked wire), that different surgeons did the insertions, and that the second experiment was separated from the first by over a year. The basic question then seems to be answered: EMG measurements are repeatable. It is at least a possibility, then, that some of the contradictory results found by different investigators can be attributed to intersubject variability and not necessarily to variations in data recording techniques.

REFERENCES

Basmajian, J. and G. Stecko. (1962) A new bipolar indwelling electrode for electromyography. J. Appl. Physiol. 17, 849.

Cooper, F.S. (1965) Research techniques and instrumentation: EMG. Proc. Conf.: Communicative Problems in Cleft Palate, ASHA Reports, No. 1. 153-168.

Faaborg-Andersen, K. (1957) Electromyographic investigation of intrinsic laryngeal muscles in humans. Acta Physiol. Scand. 41, Suppl 140.

Faaborg-Andersen, K. (1965) Electromyography of Laryngeal Muscles in Humans: Techniques and Results. Basel: S. Karger.

Gay, T. and K.S. Harris. (In press) Some recent developments in the use of electromyography in speech research. JSHR 14.

Hirano, M. and J. Ohala. (1969) Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. JSHR 12, 362-373.

Hirano, M., J. Ohala, and W. Vennard. (1969) The function of the laryngeal muscles in regulating fundamental frequency and intensity of phonation. JSHR 12, 616-628.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Port, D. (1971) The EMG data system. (See this Status Report.)
Sawashima, M., M. Sato, S. Funasaka, et al. (1959) Electromyographic study of the human larynx and its clinical application. Jap. J. Otol. Tokyo 61, 1357-1364.
Sawashima, M., T. Gay, and K.S. Harris. (1969) Laryngeal muscle activity during vocal pitch and intensity changes. Haskins Laboratories Status Report on Speech Research SR-19/20, 211-220.

An Electromyographic Study of Laryngeal Adjustments During Speech Articulation: A Preliminary Report

Hajime Hirose*
Haskins Laboratories, New Haven

INTRODUCTION

The aim of the present study is to examine the electromyographic activities of the intrinsic laryngeal muscles during speech articulation. Electromyographic (EMG) study of the laryngeal muscles in relation to the laryngeal articulatory mechanism is, at present, still in the preliminary stages mainly because of technical difficulties in obtaining reliable data without disturbing the natural movements of the articulatory organs. In the present study, an attempt was made to insert hooked-wire electrodes into the posterior cricoarytenoid and the interarytenoid perorally by indirect laryngoscopy in order to achieve accurate electrode placement while preserving natural articulatory activity. Percutaneous insertion, similar to that previously described in the literature (Hirano and Ohala, 1969; Hirose, in press), was used for EMG recordings from the rest of the laryngeal muscles. The activities of five intrinsic laryngeal muscles were, thus, systematically examined with special reference to the articulation of American English.

EXPERIMENTAL PROCEDURES

EMG recordings were made using hooked-wire electrodes. In order to insert the electrodes into the posterior cricoarytenoid (PCA) and the interarytenoid (INT), a peroral approach was attempted using a specially designed needle holder which permitted insertion of the electrodes into the target muscles by indirect laryngoscopy under topical anesthesia. Insertion of the electrodes into the cricothyroid (CT), the thyroarytenoid (VOC), and the lateral cricoarytenoid (LCA) was performed percutaneously. More detailed description of the insertion techniques, including preparation of the electrodes and the route of insertion, as well as a description of the data-recording system and computer processing, are elsewhere in this Status Report (Hirose, 1971; Port, 1971).

The present experiment was performed on two subjects, both native speakers of American English; for one subject, two separate recordings were made, thus giving three sets of final data. The subjects were required to read randomized lists of stimulus words ten to sixteen times each. Table I lists the muscles examined and the types of stimulus words used in each experiment.

---

*Also, Faculty of Medicine, University of Tokyo.

TABLE I

MUSCLES EXAMINED*

Subject 1 (LL)	Subject 2 (LJR) Series A	Subject 2 (LJR) Series B
posterior crico-arytenoid (PCA)	PCA	PCA
interarytenoid (INT)	VOC	lateral cricoarytenoid (LCA)
thyroarytenoid (VOC)	CT	OO
cricothyroid (CT)	sternohyoid (SH)	sternothyroid (ST)
	orbicularis oris (OO)	superior constrictor (SC)
		genioglossus (GG)
		geniohyoid (GH)
		mylohyoid (MH)

*For Subject 1 and for series A of Subject 2, an attempt was made to record from five intrinsic muscles simultaneously. However, unsatisfactory recordings were obtained for the LCA of Subject 1 and for the LCA and INT of Subject 2. For series B of Subject 2, only the LCA and the PCA were selected as representatives of the intrinsic laryngeal muscles.

TYPES OF STIMULUS WORDS

Subject 1 (LL)	Subject 2 (LJR) Series A	Subject 2 (LJR) Series B
əC̲ʌp C: p,b,s,z,h	əC̲ʌp C: p,b,t,d,k,g,f̲,v,s, z,θ,ð̲,ʃ,ʒ,č̲,j̲,h	əC̲ip C: b,s,z,t,d,h
bʌC̲ə C: p,b,s,z,h	bʌC̲ə C: as above	əpV̲C V: ɪ,i,ʊ,u,ɛ,e, C: p,b,t,d,k,g,s,z
əbʌC̲ C: p,b,s,z,h	əbʌC̲ C: as above	meaningful words:
ʌbəC̲ C: p,b,s,z	ʌbəC̲ C: as above	a pit, a sit, a bit, a fit, a spit, a cap, a split, a gap, a hit, a fan, a van
pʌpə	pʌp ə	
hʌpə	hʌp ə	
number of stimulus types = 20	number of stimulus types = 69	number of stimulus types = 77

RESULTS

Each of the EMG curves in Figures 1 through 3 represents a computer average of ten to sixteen utterances. The line-up point (0 on the time axis) in these samples was selected at the voice offset of the stressed vowel before the final stop consonant. The timing marks for the acoustic signals in each figure were obtained by averaging values measured from oscillographic records.

Posterior Cricoarytenoid (PCA)

The general EMG pattern of the PCA clearly demonstrates the voiced/voiceless contrast. In the case of [əpʌp] in Figure 1, for example, the EMG activity of the PCA starts to decrease approximately 250 msec prior to the onset of initial [ə]. The activity then begins to increase 100 msec before the stop closure of [p], reaching the peak 110 msec prior to the stop release, and immediately begins to decrease with the production of the stress vowel [ʌ]. Approximately 110 msec prior to the voice offset, it shows a steep rise again for the final [p].

In [əbʌp], on the other hand, the PCA activity stays low throughout the voiced period from the initial vowel to the stressed vowel, including intervocalic [b]. It should be noted, however, that the EMG curve ascends slightly 110 msec prior to the release of [b], then descends again approximately at the time of the release, and finally rises steeply starting 40-50 msec before the voice offset.

The general patterns of PCA activity described above are also found in Figures 2 and 3.[1]

Interarytenoid (INT)

The INT showed a reciprocal pattern of activity when compared to the PCA in relation to the voiced/voiceless contrast.

As illustrated in Figure 1, INT activity in the case of [əpʌp] begins to increase 250 msec prior to the initial vowel production, reaching its peak when PCA activity reaches its valley. In general, the INT shows a sort of inversion of the pattern of PCA activity throughout the utterance.

For the articulation of [əbʌp], the INT shows more or less continuous activity for the voiced segments after the initial rise in activity, but there is some decrease in activity for intervocalic [b] as compared to the neighboring vowel segments.[2]

---

[1] The question of differences in the amplitude or the duration of the averaged EMG activity for the same phoneme with respect to phoneme environment will not be considered in this paper.

[2] The tendency of INT activity to be lower for voiced consonants than for vowel segments was more clearly revealed in the case of voiced fricatives, whose data are not shown here.

Superimposed Averaged EMG Signals of the Intrinsic Laryngeal Muscles of Subject 1 for the Utterances [əpʌp] and [əbʌp]

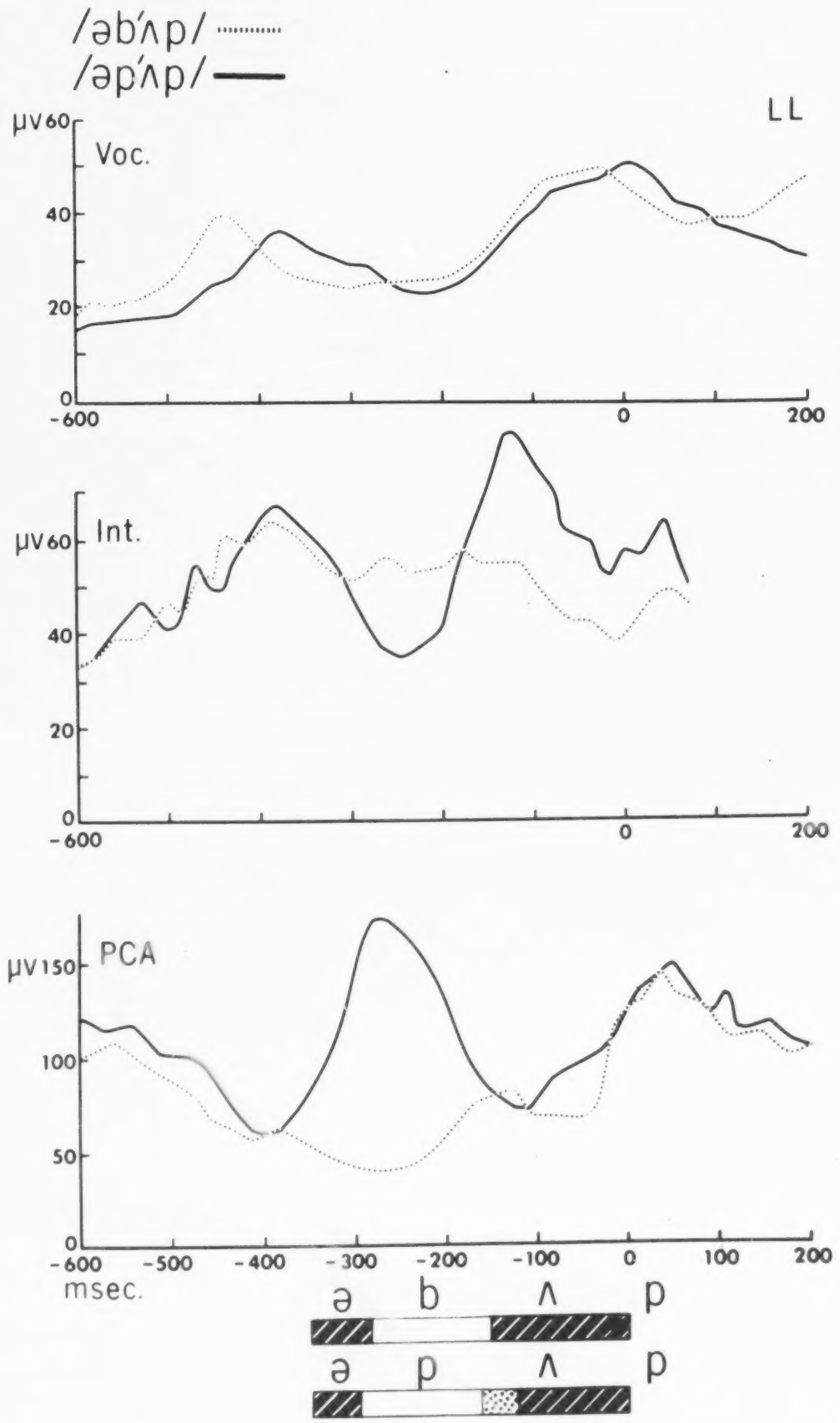


Fig. 1

Note: From top to bottom, traces represent the signals of the thyroarytenoid, the interarytenoid, and the posterior cricoarytenoid. The line-up point (0 on the abscissa) indicates the voice offset of the stressed vowel before the final stop closure.

Superimposed Averaged EMG Signals of, from Top to Bottom, the Orbicularis Oris, the Posterior Cricoarytenoid, the Cricothyroid, and the Thyroarytenoid of Subject 2, Series A, for the Utterances [əpʌp] and [əbʌp]

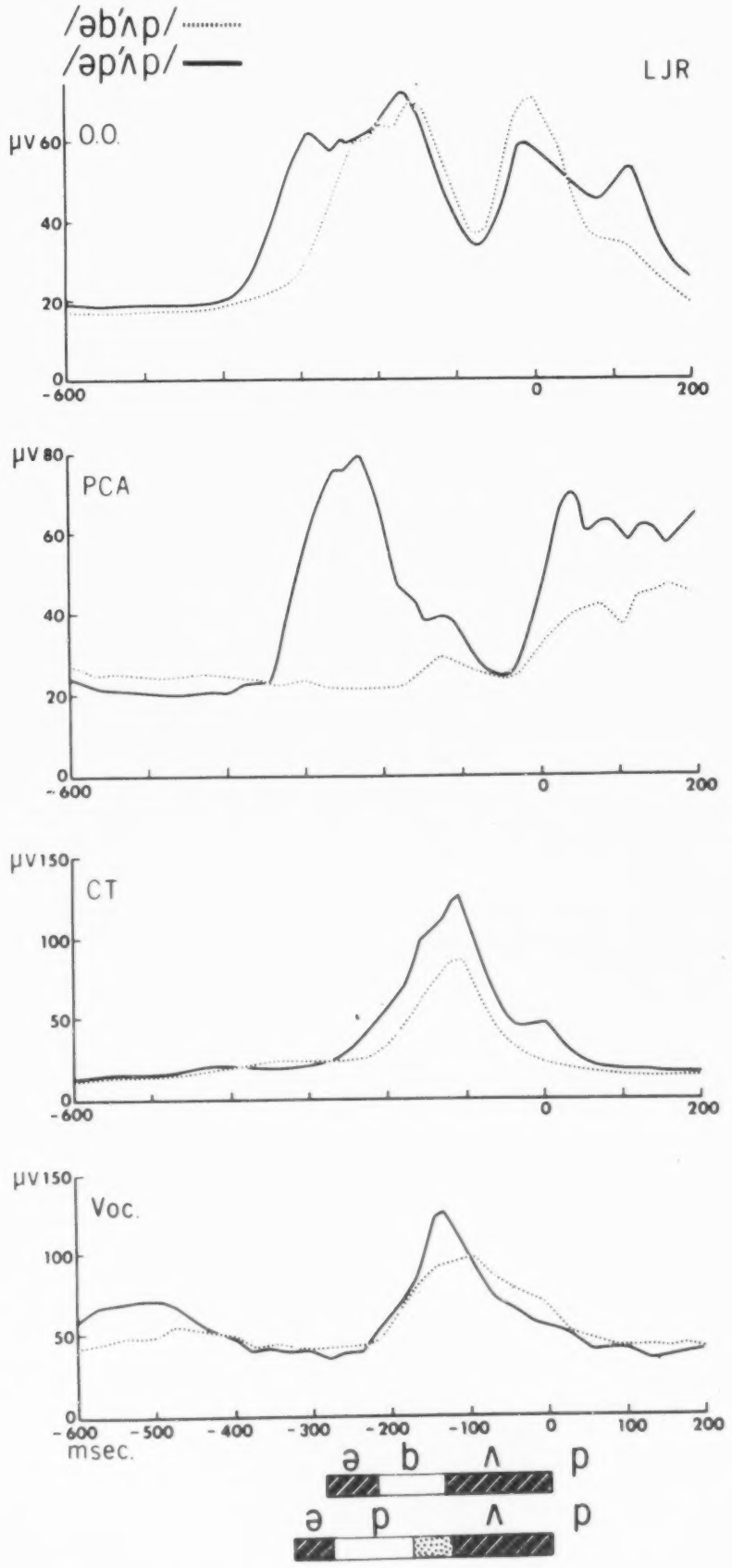


Fig. 2

Superimposed Averaged EMG Signals of, from Top to Bottom, the Orbicularis Oris, the Sternohyoid, the Lateral Cricoarytenoid, the Posterior Cricoarytenoid, and the Superior Constrictor of Subject 2, Series B for the Utterances [əpɪt] and [əbɪt]

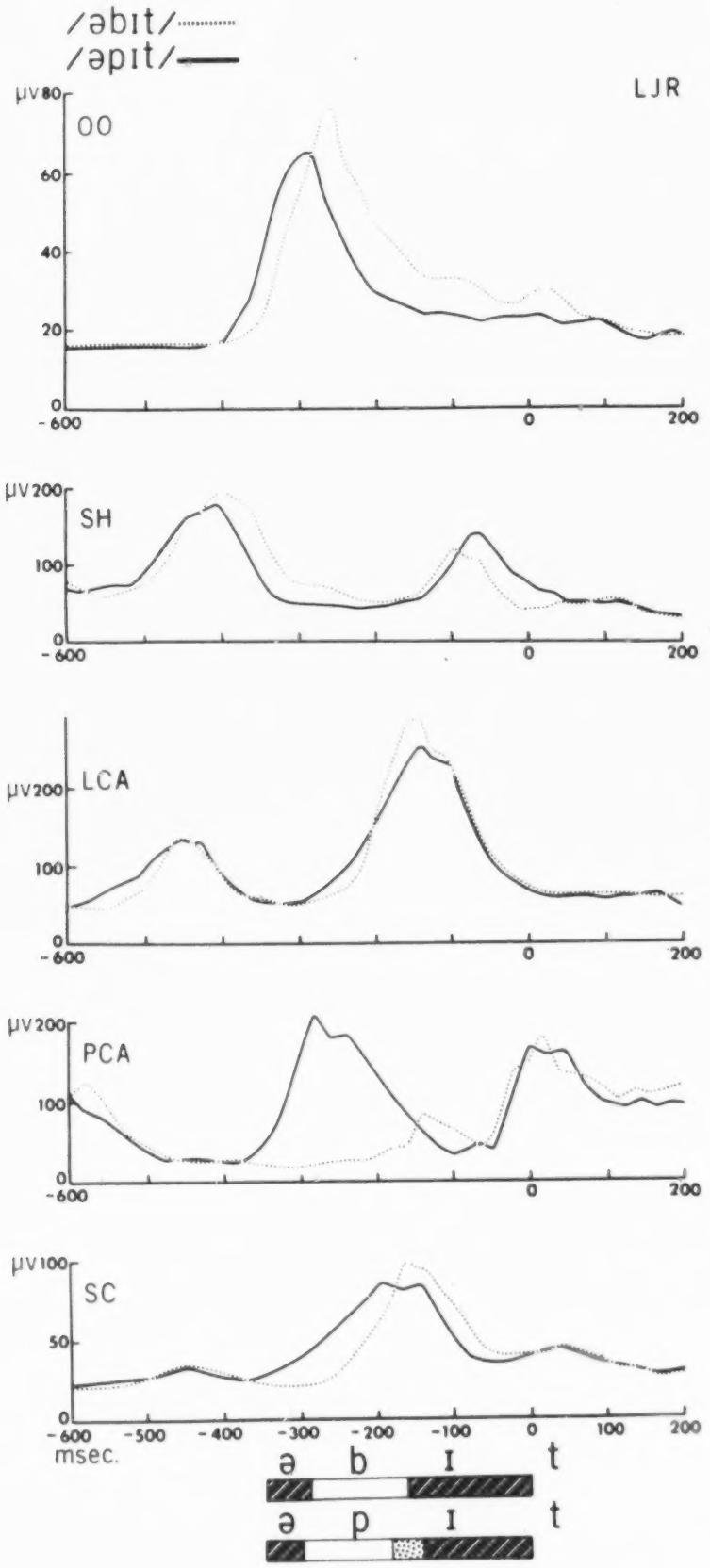


Fig. 3

INT activity for the stressed vowel [ʌ] is higher after voiceless [p] than after voiced [b] as illustrated in Figure 1.

Thyroarytenoid (VOC) and Lateral Cricoarytenoid (LCA)

The general patterns of EMG activity of the VOC (Figures 1 and 2) and the LCA (Figure 3) are different from those of the PCA or the INT with regard to the voiced/voiceless contrast. The most consistent findings on these two so-called adductors are that EMG activity decreases for the consonant segments regardless of the voiced/voiceless distinction and that the activity shows a definite increase for vowel segments, particularly for the initiation of voicing and for stressed vowels.

Cricothyroid (CT)

The CT (Figure 2) shows a temporary increase in EMG activity for the stressed vowel but presents no consistent differences in relation to the voiced/voiceless contrast.[3]

EMG Activity of the Other Articulatory Muscles

The orbicularis oris (OO) showed increasing activity for [p] and [b] articulation, as we would expect from findings of previous studies (Fromkin, 1966; Harris et al., 1965; Tatham and Morton, 1969). For [p], its activity starts to increase synchronously with, or 30-40 msec before, the increase in PCA activity (Figures 2 and 3). The duration of its activity for intervocalic [p] varies from that of the PCA.

EMG activity of the superior constrictor and the sternohyoid, included in Figure 3, will not be discussed here.

COMMENT

Since the introduction of the use of hooked-wire electrodes which is usually combined with the percutaneous insertion technique, a considerable number of reports have accumulated on laryngeal muscle activity during speech and singing. In general, these studies, with related anatomical and modeling work, support the classical division of the intrinsic laryngeal muscles into three functional groups--abductor (PCA), tensor (CT), and adductor (INT, LCA, VOC[4]). However, little attempt has been made to clarify the function of the laryngeal muscles in consonant articulation.

In particular, participation of the PCA in speech has not been systematically studied, although the function of the PCA as a respiratory muscle has been

---

[3]The peak amplitude of the CT is apparently higher for the stressed vowel following the voiceless consonant [p] than for that following the voiced [b]. The difference is, however, not consistent for other sets containing a voiced/voiceless contrast, whose data are not shown here.

[4]The thyroarytenoid is generally believed to have an adducting effect, as well as a shortening and tensing effect, on the vocal fold.

well documented (Pressman, 1942; Suzuki and Kirchner, 1969). As far as PCA activity in phonation is concerned, Faaborg-Andersen (1957) reported that EMG activity of the PCA decreases during sustained phonation. Kotby and Haugen (1970), on the other hand, observed increasing activity of the PCA during phonation and postulated that the PCA is not solely an abductor muscle. Dedo (1970) also reported increasing activity in the PCA during phonation in some of his clinical cases. The data of these authors are concerned exclusively with sustained vowel phonation, the fundamental frequency of which was not definitely specified.

Hiroto et al. (1967) examined laryngeal muscle activity for some Japanese words containing an intervocalic fricative [s] and stated that there was a temporary change in the electrical activity of all the intrinsic laryngeal muscles, except for the cricothyroid, corresponding to voiceless consonant articulation. What they observed in their data was an apparent increase in PCA activity accompanied by a decrease in the activity of the adductors for articulation of the intervocalis [s]. Hirano and Ohala (1969) showed one example of a raw EMG record of the PCA, illustrating increasing activity for release of glottal stops with reciprocally decreasing activity in the INT.

In the present study, it was clearly revealed that the PCA actively participates in the laryngeal articulatory adjustments, particularly for the voiced/voiceless distinction. There is a consistent increase in PCA activity for voiceless consonant production regardless of the difference in phonetic environment.[5]

In addition to gross adjustment of the glottal condition, as in voiceless consonant articulation, the PCA also appears to participate in finer adjustments, as is seen near the end of the [b] segment in Figures 1, 2, and 3 where the glottis seems to be slightly opened by minor PCA activity to permit a possible escape of air through the narrowed glottis.

The laryngeal gestures necessary for consonant production should require rapid muscle adjustments in both the abductor and the adductor groups of the larynx. Although there is some controversy about the contraction properties of the PCA of experimental animals (Hast, 1967; Hirose et al., 1969; Mårtensson and Skoglund, 1964), the present data suggest that the human PCA is able to execute fast contraction equivalent to that of the adductor muscles in laryngeal articulatory adjustments.

In a study of EMG activities of the laryngeal muscles in singing (Gay et al., 1971), we observed that PCA activity is generally suppressed during sustained phonation except for the very end of voicing in low and medium frequency ranges in the chest register and in falsetto, while it increases for high chest voice phonation. The increasing PCA activity in the latter condition may reflect the counterbalancing function of the abductor for the strong contraction of the adductors, as suggested in previous literature (Pressman, 1942; Suzuki and Kirchner, 1969). Another possibility is that different kinds of motor units are participating in the execution of muscle contraction in different conditions of phonation, since there is evidence, at least in animal experiments, that the PCA consists of several kinds of motor units (Suzuki and Kirchner, 1969).

---

[5] The increase in PCA activity for voiceless consonants is not so marked when the voiceless consonant is in absolute initial position, as in [pʌpə]. Such data are not shown here.

Although the function of the PCA, particularly in sustained phonation, should be a subject for further investigation, the role of the PCA as an abductor in speech articulation is demonstrated in the present study.

The present data indicate that there is apparently reciprocal activity between the PCA and the INT. In this sense, the INT can be considered to be a representative adductor of the vocal fold. As described above, there is an apparent difference in the degree of INT activity for vowel segments depending on the preceding consonant. Since the EMG activity represents the muscle action necessary for obtaining effective force or displacement, the degree of the activity should be higher if, for example, the displacement is larger. The glottal width is obviously larger in the articulation of voiceless consonants than in the articulation of voiced consonants, as observed in recent fiberoptic studies (Sawashima, 1968, 1970). Therefore, it is reasonable that the activity of the INT, which is responsible for adducting the vocal fold, should be greater after voiceless consonants accompanying the more widely abducted glottis.

The fact that INT activity is apparently less for voiced consonants than for vowels indicates that there is a difference in laryngeal adjustment between vowel and voiced consonant.

It should be noted that the other laryngeal adductors, the LCA and the VOC, appeared to be activated only for vowel production. It is possible that the LCA and the VOC are not merely adductors of the vocal fold and that there is functional differentiation within the group of so-called adductor muscles. The lesser activity of the LCA and the VOC for the production of voiced consonants than for vowels again suggests a difference in glottal adjustments between these two phonetic conditions. If these adductors are less active for voiced consonant production than for vowel production, the glottal closure, and possibly the tension of the vocal folds, should tend to be less in the former condition. Table II illustrates a combination of the activities of the functionally different laryngeal muscle groups as a possible physiological correlate for different phonetic conditions.

TABLE II

	Abductor	Adductor	
	PCA	INT	VOC LCA
vowel	-	++*	+~++*
voiced consonant	-	+	-
voiceless consonant	+	-	-

*++ represents high activity, while + indicates moderate activity. In the INT, for example, EMG activity was evident both for the vowel and for the voiced consonant but more so for the vowel. In the VOC and the LCA, the degree of EMG activity for vowel production apparently differs depending on the prosodic condition, i.e., it is higher (++) for stressed vowels.

The present report is based on the analysis of the limited amount of data processed thus far. A more detailed report will follow with reference to laryngeal articulatory adjustments. The effects of phonetic environment and phonetic categories on laryngeal muscle activity will be further considered.

REFERENCES

Dedo, H. (1970) The paralyzed larynx: An electromyographic study in dogs and humans. Laryngoscope 80, 1455-1517.

Faaborg-Andersen, K. (1957) Electromyographic investigation of intrinsic laryngeal muscles in humans. Acta physiol. Scand. 41, Suppl. 140.

Fromkin, V. (1966) Neuro-muscular specification of linguistic units. Lang. Speech 9, 170-199.

Gay, T., H. Hirose, M. Strome and M. Sawashima. (1971) Electromyography of the intrinsic laryngeal muscles in singing. (See this Status Report.)

Harris, K., G. Lysaught, and M. Schvey. (1965) Some aspects of the production of oral and nasal labial stops. Lang. Speech 8, 135-147.

Hast, M.H. (1967) The respiratory muscle of the larynx. Ann. Otol. Rhinol. Laryng. 76, 489-497.

Hirano, M. and J. Ohala (1969) Use of the hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. JSHR 12, 362-373.

Hirose, H., T. Ushijima, T. Kobayashi and M. Sawashima. (1969) An experimental study of the contraction properties of the laryngeal muscles in the cat. Ann. Otol. Rhinol. Laryng. 78, 297-307.

Hirose, H. (In press) The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. To appear in Phonetica.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Hiroto, I., M. Hirano, Y. Toyozumi and T. Shin. (1967) Electromyographic investigation of the intrinsic laryngeal muscles related to speech sounds. Ann. Otol. Rhinol. Laryng. 76, 861-872.

Kotby, M.N. and L.K. Haugen. (1970) Critical evaluation of the action of the posterior crico-arytenoid muscle, utilizing direct EMG-study. Acta Otolaryng. 70, 260-268.

Mårtensson, A. and C.R. Skoglund. (1964) Contraction properties of intrinsic laryngeal muscles. Acta physiol. Scand. 60, 318-336.

Port, D.K. (1971) The EMG data system. (See this Status Report.)

Pressman, J.J. (1942) Physiology of the vocal cords in phonation and respiration. Arch. otolaryng. 35, 335-398.

Sawashima, M. (1968) Movements of the larynx in articulation of Japanese consonant. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo) No. 2, 11-20.

Sawashima, M. (1970) Glottal adjustments for English obstruents. Status Report on Speech Research (Haskins Labs.) SR 21/22 Jan-June, 187-200.

Suzuki, M. and J.A. Kirchner. (1969) The posterior cricoarytenoid as an inspiratory muscle. Ann. Otol. Rhinol. Laryng. 78, 849-864.

Tatham, M.A.A. and K. Morton. (1969) Some electromyography data toward a model of speech production. Lang. Speech 12, 39-53.

The Velopharyngeal Mechanism: An Electromyographic Study
A Preliminary Report

Fredericka B. Berti*
Haskins Laboratories, New Haven

The velopharynx has been of interest to those concerned with the clinical problems of anomalous mechanisms (i.e., cleft and nonfunctional palates) for some time. More recently, linguists have become interested in the velopharynx as the possible site of certain distinctive phonological features. This report presents some of the preliminary findings of an electromyographic (EMG) study designed to describe the activity patterns of the velopharyngeal mechanism in 1) oralization-nasalization gestures, 2) voicing distinctions, and 3) adjustment of pharyngeal cavity size.

Fritzell (1969) surveyed and expanded our knowledge of the velopharynx. He reports that the levator palatini and the superior pharyngeal constrictor present an "on-off" pattern of activity corresponding to oral and nasal sounds (p. 31), with greater levator activity for high than for low vowels (pp. 47, 63). Lubker (1968) also reports greater EMG activity for high vowels than for low vowels and greater velar elevation, measured on cinefluorographic films, for high vowels than for low vowels. Lubker et al. (1970) also report higher peak EMG potentials in the levator for voiceless consonants than for the cognate voiced consonants. Fritzell reports, too, that the palatopharyngeus shows great intersubject variation and no consistent pattern of oralization or nasalization activity but demonstrates its greatest activity for low vowels. The palatoglossus is there reported to be active in lowering the velum for the production of nasals and also in raising the mid and dorsal portions of the tongue in the production of velar phonemes. Hence, / ŋ / shows greater potentials than /g/, as the palatoglossus is performing two functions during / ŋ / but only one for /g/ (p. 48). In addition, Fritzell reports greater palatoglossus activity for /u/ than for / ɑ / or /i/. Moll (1962) and Moll and Shriner (1967) made cinefluorographic measurements of the velopharynx during speech and hypothesize that velar height is greater for high vowels than for low vowels as a consequence of the connection of the soft palate to the tongue by the palatoglossus, limiting the distance between the dorsum of the tongue and the velum to the maximum length of the palatoglossus muscle. Lubker et al. (1970) report palatoglossus activity corresponding to nasal consonant production and to three instances of nonnasal articulations: /u/, whose peak height exceeded all but one nasal-associated peak, as well as two peaks associated with word-initial oralization gestures. In addition, these same authors found small bursts of palatoglossus activity accompanying strong levator activity.

*Also The Graduate Center of the City University of New York.

Perkell (1969) and Kent and Moll (1969) report measurements of the velopharynx in two cineradiographic studies. Both studies show larger velopharyngeal cavity sizes during voiced stop consonant production than during the production of the homorganic voiceless stops. Perkell attributes this difference to the tension difference postulated by Chomsky and Halle (1968) as a distinctive feature separating the two groups of stop consonants, /p,t,k/ and /b,d,g/. According to this theory, the superior pharyngeal constrictor contracts for the production of /p,t,k/, thus reducing pharyngeal cavity size and increasing intra-oral air pressure, thereby reducing the transglottal pressure drop necessary for the maintenance of glottal pulsing. Kent and Moll, however, feel that the enlargement is not purely passive, but rather that there is active enlargement of the velopharynx, perhaps by movement of the hyoid bone and associated structures. They proposed that the sternohyoid might have this function.

EXPERIMENTAL PROCEDURE

In the present study, bipolar hooked-wire electrodes were inserted into the dimple of the levator palatini, the superior pharyngeal constrictor, the middle pharyngeal constrictor, the palatopharyngeus, and the palatoglossus. All of these insertions were peroral. In addition, percutaneous insertions were used for the placement of electrodes in the sternohyoid and the orbicularis oris. The EMG potentials were recorded, along with the audio signal and automatic timing markers, onto magnetic tape. The potentials were rectified, integrated, and computer averaged. More detailed descriptions of the insertion techniques are found in Hirose (1971); the data processing system is described by Port (1971).

There were two sets of stimuli:

1. C_1VC_2, where V is /i,u, ɑ, æ / and where $C_1 = C_2$ and they are either /p/ or /b/;

2. $fV_1C_1C_2V_2p$, where $V_1 = V_2$ and they are /i,u, ɑ/ and where C_1 and C_2 are different, C_1 being /m,ŋ,p,t,k,b,d,g/ and C_2 being /m,p,t,k,b,d,g/.

These subjects were three adult native speakers of various dialects of American English.

RESULTS

The EMG data curves presented in Figures 1-7 represent averages of nine to sixteen tokens of each stimulus type. Successful recordings were not achieved from all insertions in all subjects. Reports of each muscle are prefaced with the number of subjects for whom recordings were obtained. The figures presented are representative samples of the larger body of data, each subject being represented for each muscle wherever possible. The line-up point for the fVCCVp stimuli was the boundary between the consonants, and for the CVC stimuli it was the onset of voicing. The line-up point is labeled as "0" on the abscissa of Figures 1-7. Voicing onset of the first syllable of the fVCCVp stimuli and voicing offset of the second syllable of the fVCCVp and the CVC stimuli are marked with arrows in the figures.

Levator Palatini

EMG recordings from the levator palatini demonstrate distinctive patterns for the oralization and nasalization gestures. There is a burst of activity for oral speech sounds and a corresponding decrease in activity for nasal sounds. There is a peak of activity corresponding to the production of stop consonants which is greatest when the stop follows a nasal consonant (Figure 1a). Nasal coarticulation is evident in vowels immediately preceding nasal consonants, particularly when the vowel is /ɑ/ (Figure 1a).

For one subject (FBB), the levator peak is always highest for the production of a voiced stop, as compared to the peak corresponding to the voiceless homorganic stop in the same environment (Figure 1a). In the other subject for whom data are available (KSH), a distinction between voiced and voiceless homorganic stops is not evident from inspection of the averaged curves of levator activity. For both subjects levator activity increases with vowel height and backing (Figure 1b).

Middle Pharyngeal Constrictor

EMG recordings from the middle constrictor were obtained for one subject (KSH). There is a decrease in middle constrictor activity which corresponds to nasal productions (Figure 2a), but the extent of nasal coarticulation has not yet been determined. Activity in the middle constrictor is highest for /i/ and lowest for /u/ (Figure 2b).

Superior Pharyngeal Constrictor

Superior constrictor recordings were obtained from one subject (FBB). There is a decrease in superior constrictor activity which corresponds to the production of nasal phonemes. Nasal coarticulation is evident in vowels immediately preceding nasal consonants (Figure 3a). EMG activity is greater as tongue height and backing decrease (Figure 3b, c). That is, the greatest activity is found for /æ/, with peak height decreasing from /ɑ/ to /i/ to /u/. There is also a tendency for the peak EMG activity associated with stop consonant production to be greater for voiceless stops than for voiced stops in the same environments (Figures 4a, b).

Palatopharyngeus

EMG recordings of the palatopharyngeus were obtained from three subjects. There is a decrease in EMG activity associated with nasal productions (Figure 5a). Nasal coarticulation is evident in vowels immediately preceding nasal consonants (Figure 5a). In one subject (FBB), palatopharyngeus activity increases as tongue height and backing decrease (Figure 5b), so that peak levels decrease steadily from /æ/ to /ɑ/ to /i/ and finally to /u/. For subjects KSH and AA no data are available for /æ/. For these subjects, /ɑ/ demonstrates the greatest vowel activity, with /i/ and /u/ having approximately equal peak heights. Thus, patterns are similar for palatopharyngeus and superior pharyngeal constrictor. For subjects KSH and FBB peak heights generally are lower for stimuli containing voiced stops than for those containing their voiceless cognates (Figure 5a). No voicing data are yet available for subject AA.

Levator palatini

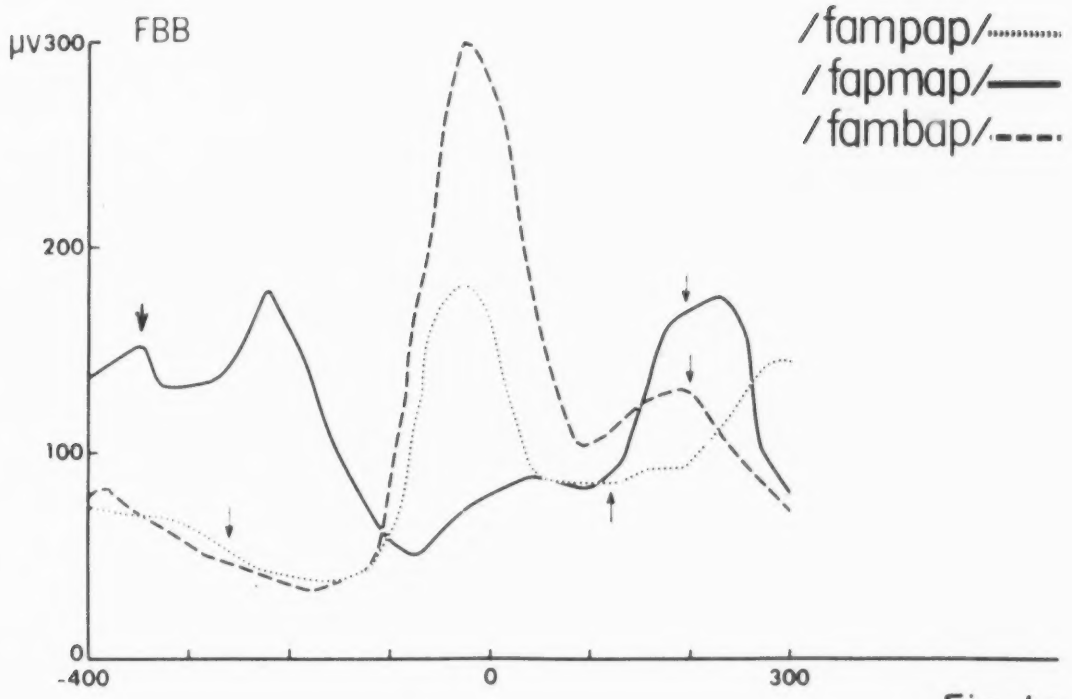


Fig. 1a

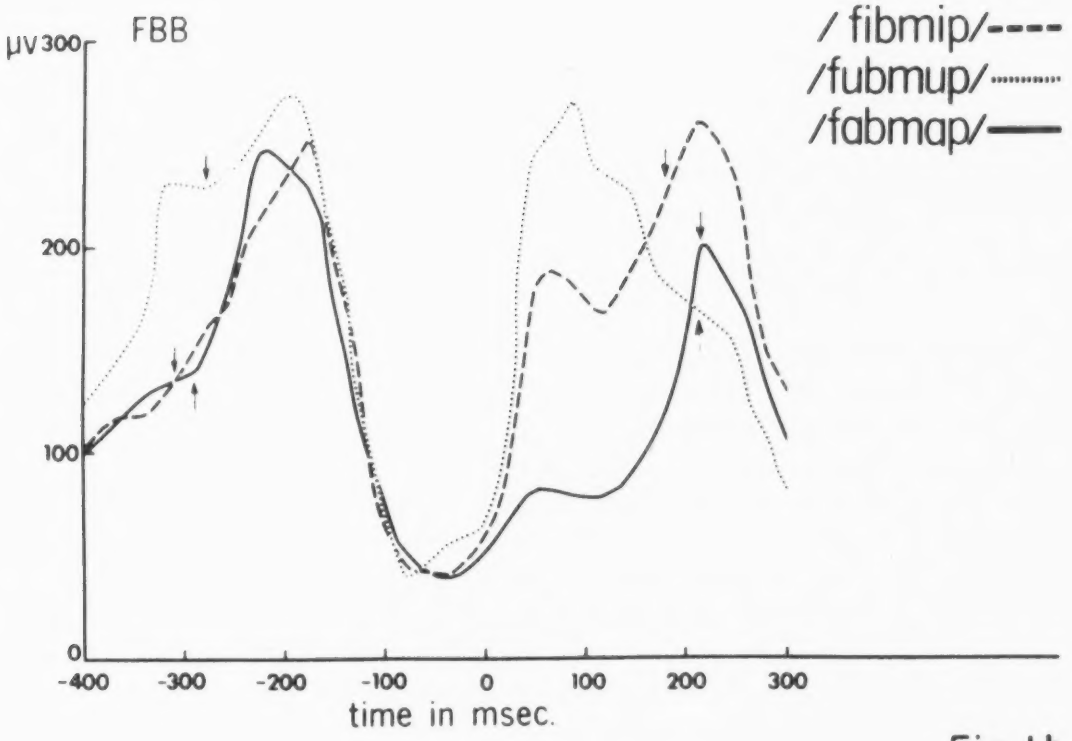


Fig. 1b

Middle Constrictor

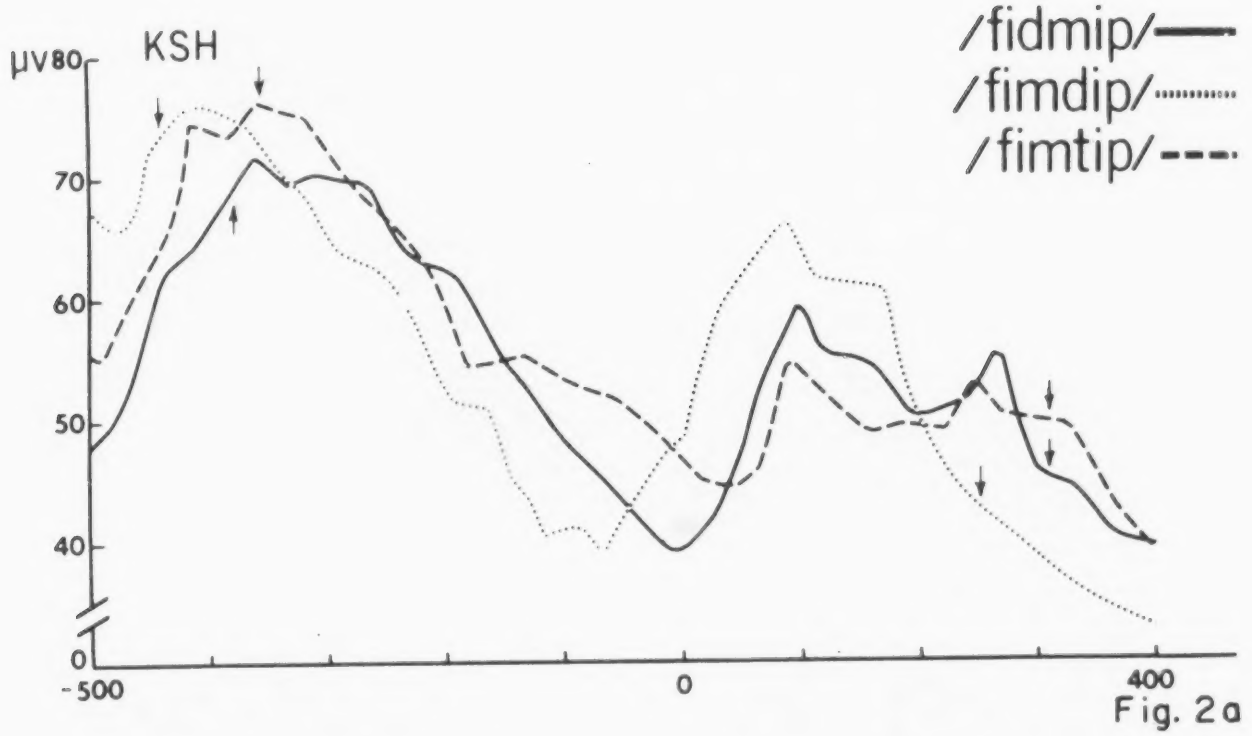


Fig. 2a

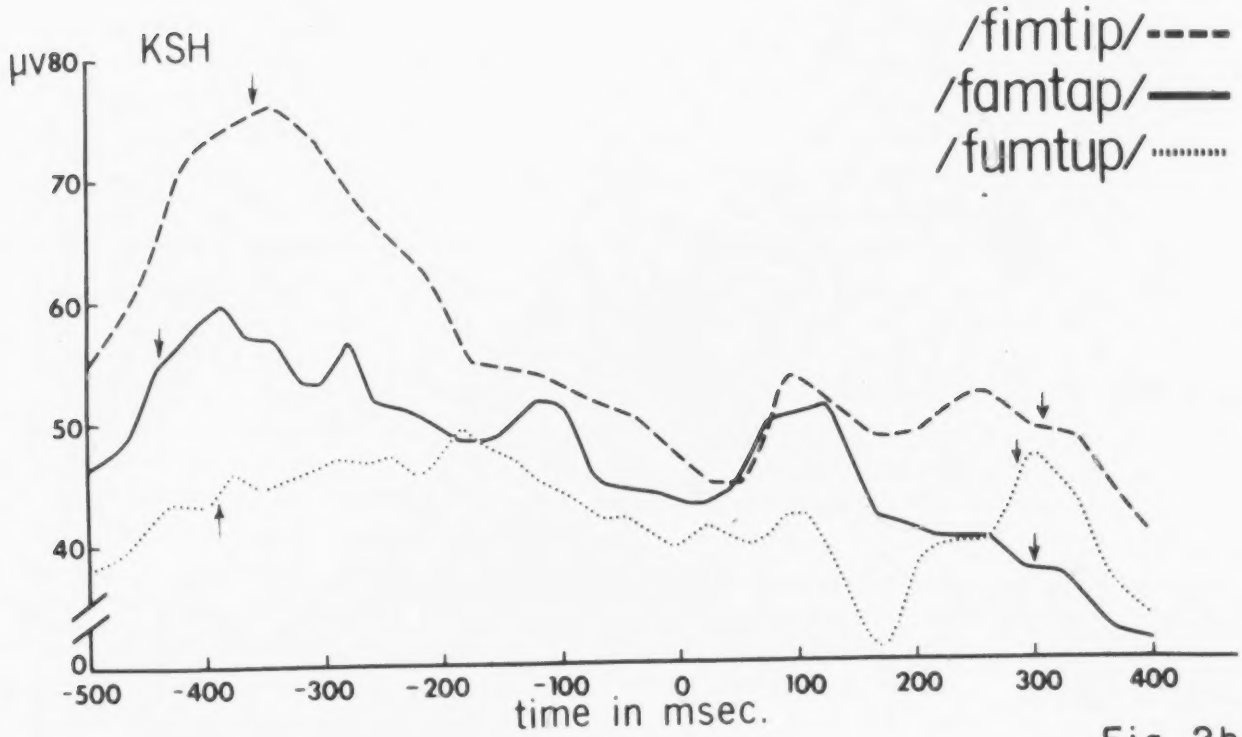


Fig. 2b

Superior Constrictor

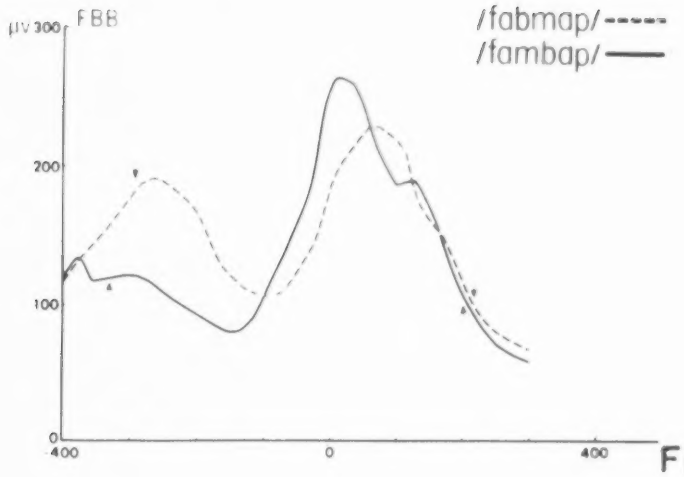


Fig. 3a

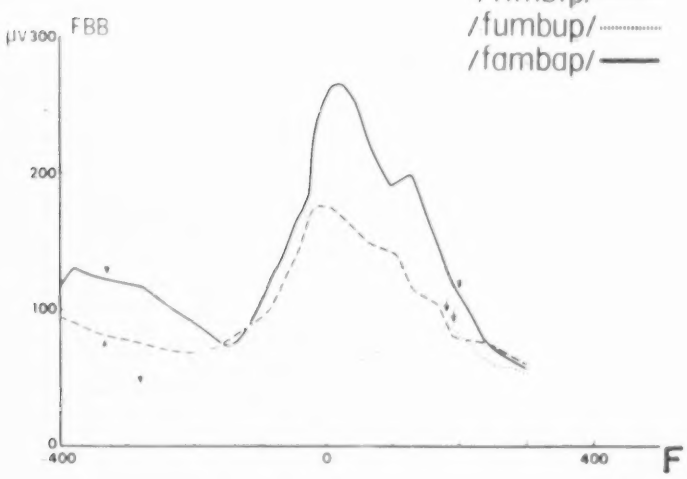


Fig. 3b

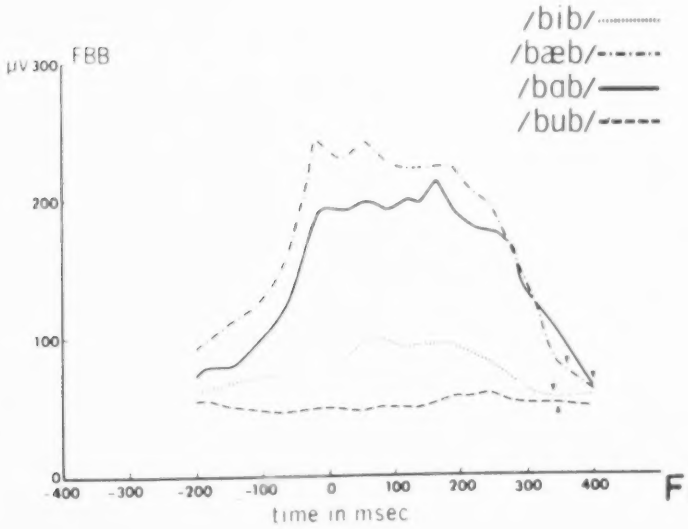


Fig. 3c

Superior Constrictor

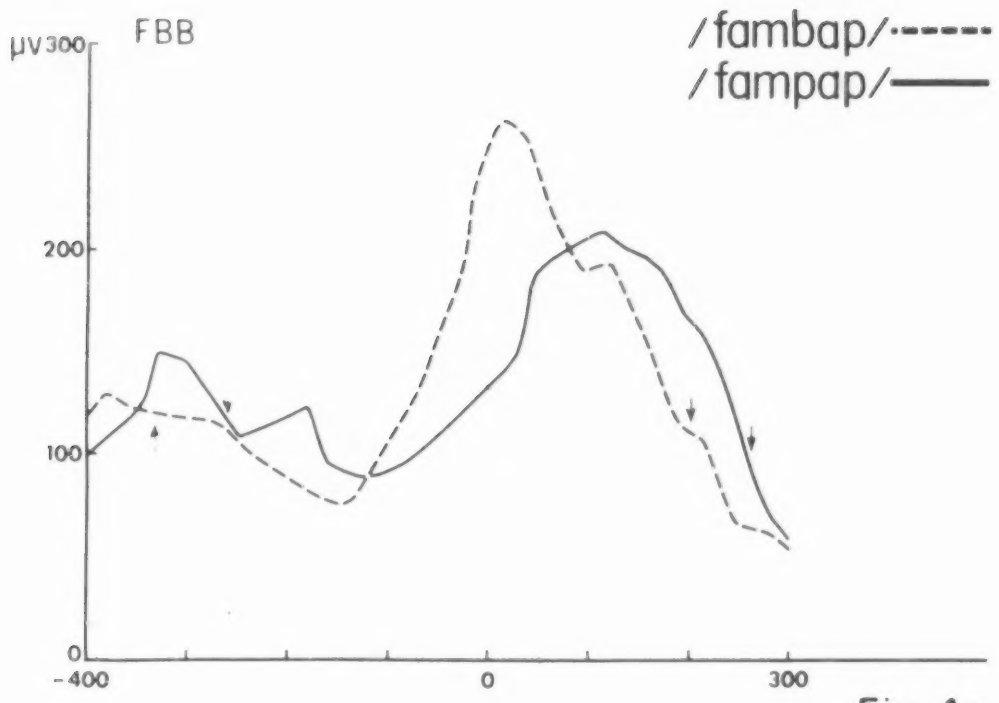


Fig. 4a

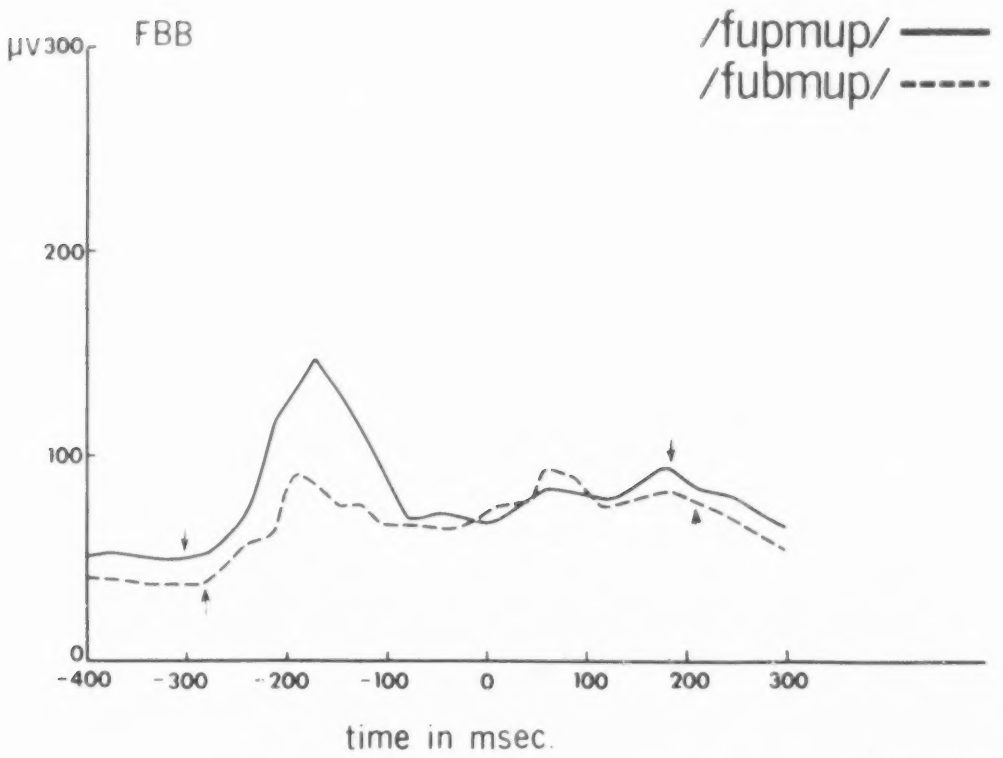


Fig. 4b

Palatopharyngeus

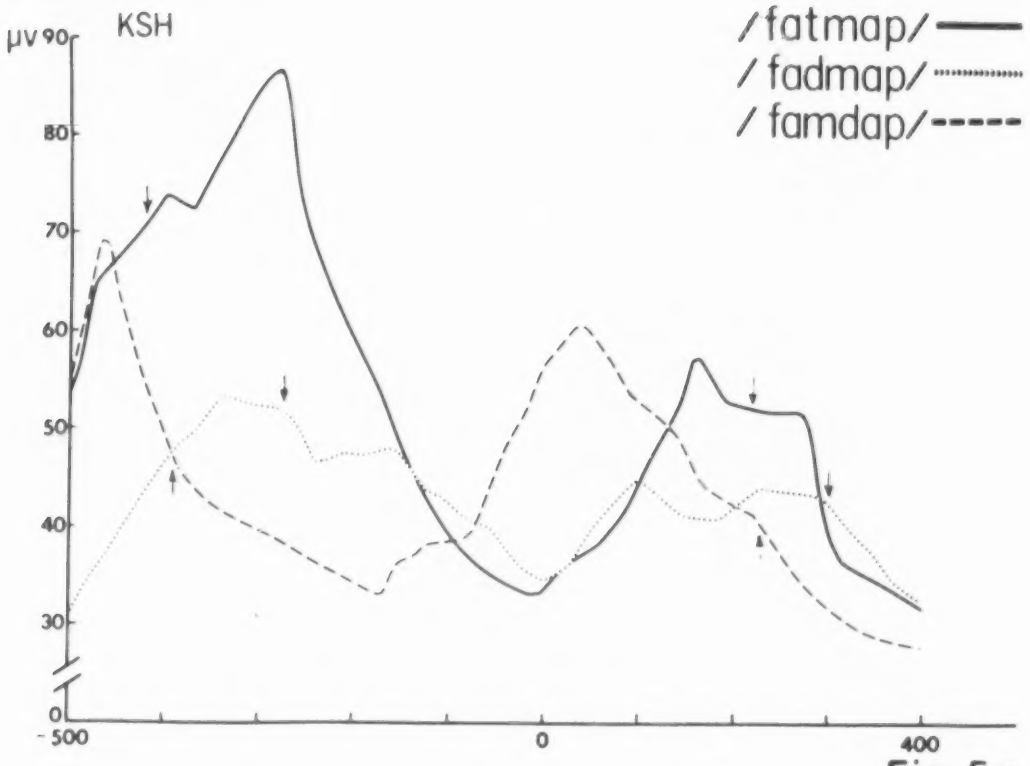


Fig. 5a

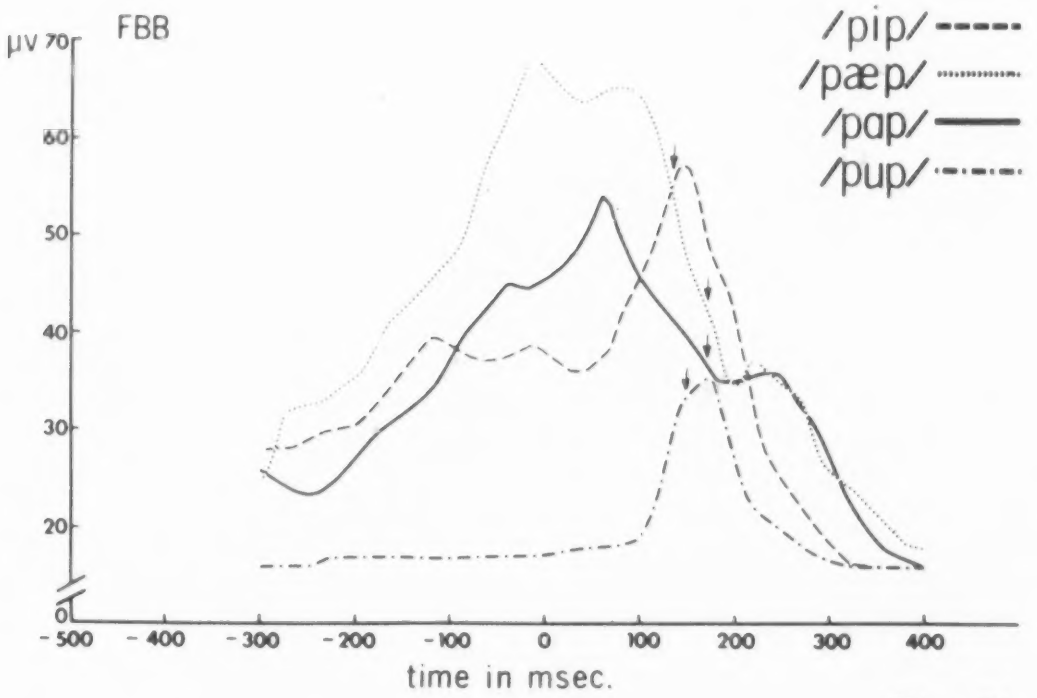


Fig. 5b

Palatoglossus

Two similar patterns of EMG activity have been found in the recordings from the palatoglossus obtained from two subjects. In one subject (FBB), activity increases with tongue backing and height, so that peaks are found for /k/, /g/, /u/, and /ɑ/, with /u/ greater than /ɑ/ (Figure 6a) and /k/ greater than /g/ (Figure 6b). Peaks are occasionally found which correspond to the production of /ŋ/ (Figure 6c), but these peaks never achieve the magnitude of those found for the corresponding oral phonemes. Otherwise, no peaks are found which correspond to the production of nasal phonemes (Figures 6a, 7a). For the other subject (AA), peaks are found which correspond to the production of /k/ and /ɑ/ (Figure 7b). Occasional peaks are found which correspond to the time of stop consonant occlusion (Figure 7c).

COMMENTS

Oralization-Nasalization Mechanism

The levator palatini is the chief effector of velopharyngeal closure, as, indeed, had long been assumed in the literature. Hence, it is active preceding any oral sound and ceases for nasal sounds.

All of the other velopharyngeal muscles under investigation in this study show a decrease in electromyographic potentials that corresponds to nasal articulation. There is no evidence that any muscle in this group acts to lower the palate for nasal coupling save the levator palatini, which accomplishes this by a decrease in activity (Figures 1a, b). The palatoglossus, which has been implicated as an active palate lowerer (Fritzell, 1969; Lubker et al., 1970), generates a pattern of activity associated with tongue backing and raising (Figures 6a, 7a,b, for example).

Voicing

Perkell (1969) has pointed to the superior pharyngeal constrictor as the site of one of the features separating the class /p,t,k/ from the class /b,d,g/, that is, as the site of the feature [±tense] postulated by Chomsky and Halle (1968). Contraction of the superior pharyngeal constrictor reduces the diameter of the pharyngeal lumen. Such a reduction in cavity size during production of a stop consonant will cause an increase in the rate of equalization of sub- and supraglottal pressures. The maintenance of a transglottal pressure differential is essential to the continuation of glottal pulsing. Hence, contraction of the superior pharyngeal constrictor will cause the cessation of glottal pulsing and result in the production of a "voiceless" or "tense" stop consonant. If the superior constrictor fails to contract, or demonstrates a decrease in electromyographic activity, pharyngeal cavity size will be greater and the transglottal pressure differential necessary to the continuation of glottal pulsing will be maintained for a longer period. This state will allow for the continuation of voicing during stop occlusion, producing a "voiced" or "lax" stop consonant.

Data collected from the superior constrictor for this study usually show lower potentials for voiced than voiceless stops, although there are some exceptions (Figures 3a,b). The palatopharyngeus appears to yield lower potentials

Palatoglossus

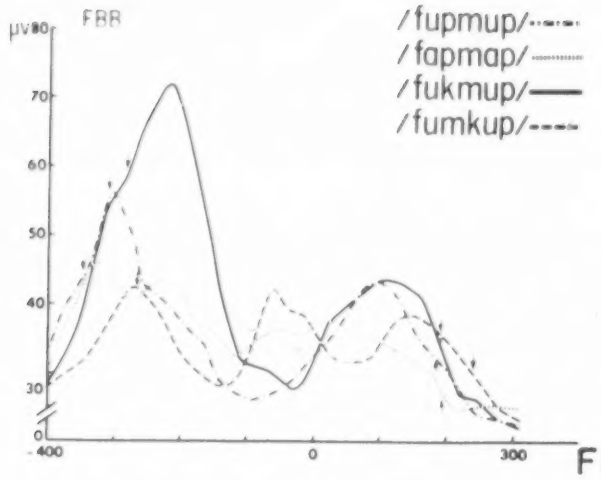


Fig. 6a

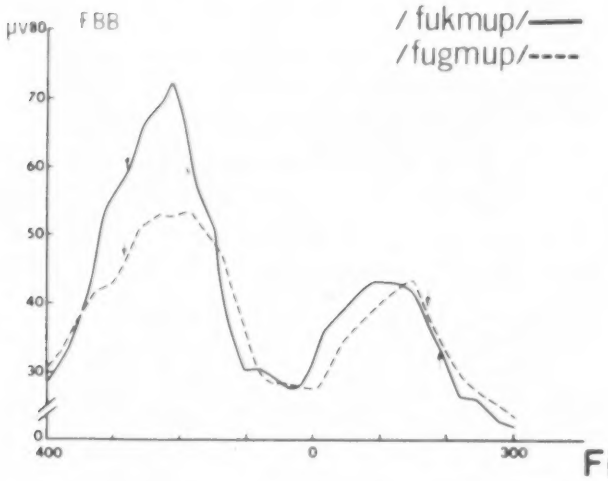


Fig. 6b

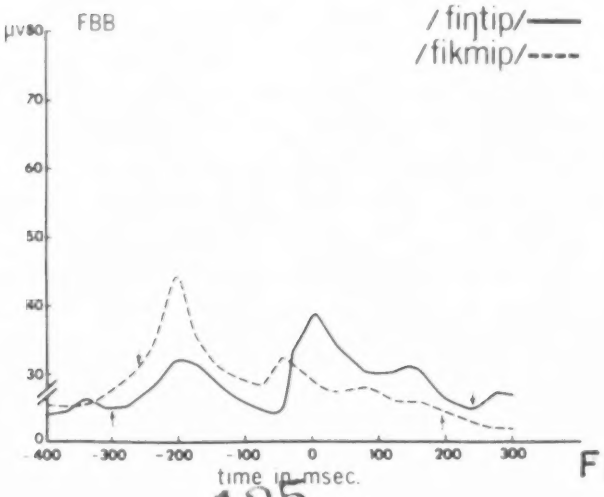


Fig. 6c

Palatoglossus

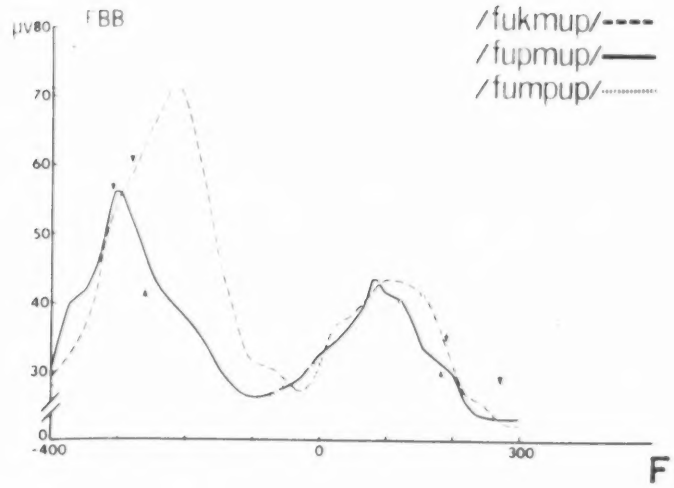


Fig. 7a

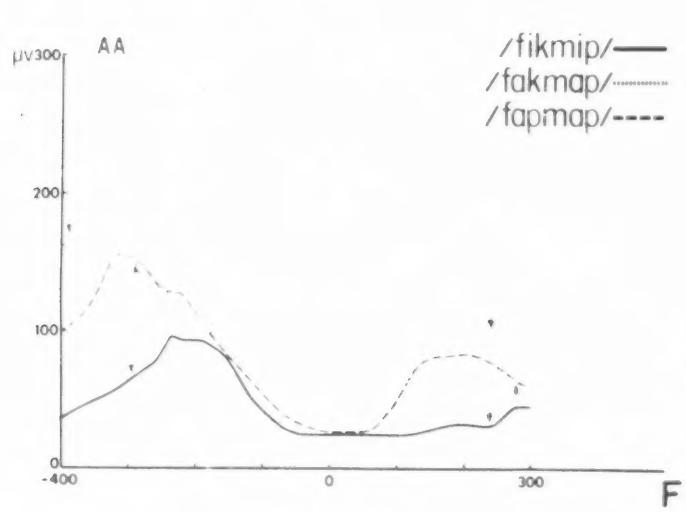


Fig. 7b

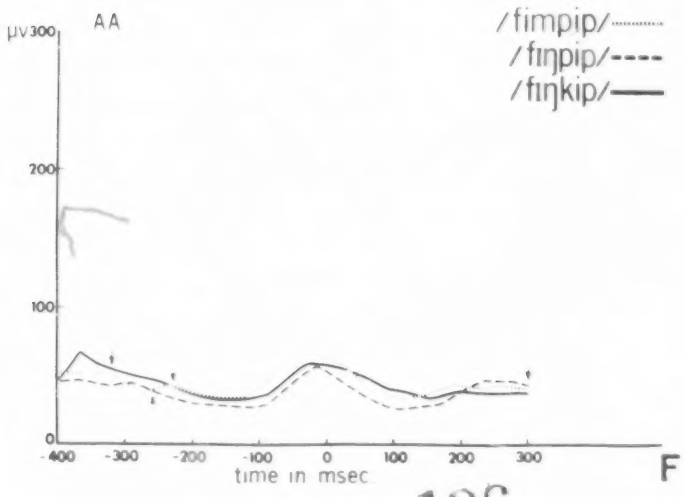


Fig. 7c

for the voiced stops than for the voiceless stops (Figure 4b). The levator palatini in one subject always yields higher potentials for the voiced stops than the voiceless stops, and in the other subject the results are variable, although the potentials corresponding to the class of voiceless stops are not consistently higher than those for the voiced stops. All of this points to a mechanism of pharyngeal enlargement which is both passive and active. Passive enlargement, noncontraction of the superior constrictor and palatopharyngeus, will result in a larger pharynegeal cavity. Active enlargement is accomplished by the increased activity of the levator palatini, whose stronger contraction will have the same effect on pharyngeal size as will noncontraction of the muscular walls of the cavity.

Aspects of Pharyngeal Cavity Size

Measurements of pharyngeal cavity width and velar height for stop consonants produced in different vowel environments show greater cavity width and velar height for stops produced in the environment of a high vowel than a low vowel (Kent and Moll, 1969). Perkell (1969) reports that the largest pharyngeal cavity for a vowel occurs during production of /u/. In this vein, we note that the levator palatini, which forms the roof of the pharyngeal cavity, demonstrates its greatest vowel potentials for /u/ and its lowest oral sound potentials for /ɑ/ (Figure 1b). Note, too, that the superior pharyngeal constrictor demonstrates its lowest potentials for /u/. Additionally, the lowest oral phoneme potentials of the middle pharyngeal constrictor correspond to /u/ production, as do those recorded from the palatopharyngeus.

The superior and middle pharyngeal constrictors and the palatopharyngeus form the walls of the pharyngeal cavity. Enlargement of the cavity in the vertical dimension may be accomplished by the elevation of the palate. Enlargement of the pharyngeal lumen may be accomplished by lessening the contractions of those muscles forming the cavity walls. Both of these patterns are found consistently in the production of high vowels, with the exception of the middle constrictor which gives its greatest peaks for the high front vowel /i/. Moll and Shriner (1967) investigated a theory which would explain the difference in velar height found for high and low vowels. They observed differences in velar elevation and postulated that they were due to the mechanical constraints of the system and not necessarily to differences in muscular force. That is, the tongue and soft palate are connected by the palatoglossus muscle, whose length determines the maximum possible distance between the velum and the dorsum of the tongue, thus requiring a lower velum for low vowels than for high vowels. This theory was modified by Lubker (1968), who showed that the observed differences in velar height reflect differences in applied muscular force necessary to achieve the degree of velopharyngeal closure required for a particular phoneme.

The current data are supportive of Lubker's (1968) modified theory, with some extension. A narrow constriction in the oral cavity, such as occurs in high vowels, will increase the impedance of the oral cavity and make it more likely that leakage through the velopharyngeal port will occur for a given velar height. If the velum is elevated further for high vowels, and the walls of the cavity remain relatively uncontracted, the impedance within the pharyngeal cavity will be decreased while the strength of the velopharyngeal seal will be increased. The result of all this will be to decrease the probability of nasal coupling during the production of high vowels. Thus, the present data indicate

a cohesive pattern of activity which adjusts pharyngeal cavity size, as a component of vowel articulation, for a constant seal.

The data and ideas presented here represent a preliminary evaluation of data collected for a larger study. There are, as noted, intersubject variations in the activity patterns of the velopharyngeal musculature (e.g., voicing distinctions in the levator palatini). The basis of these differences is as yet undetermined but may lie in differences in electrode location, dialectal variation, or idiosyncratic articulatory gestures. As was noted earlier, we have not yet achieved a complete picture of the velopharyngeal unit for all subjects. A more thorough description of the velopharyngeal mechanism, and explanation of the individual differences noted, must await further data collection and evaluation.

REFERENCES

Chomsky, Noam and Morris Halle. (1968) The Sound Pattern of English. New York: Harper and Row.

Fritzell, Bjorn. (1969) The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. Acta Otolaryngologica, Suppl. 250, 1-81.

Hirose, Hajime. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Kent, Raymond D. and Kenneth L. Moll. (1969) Vocal-tract characteristics of the stop cognates. J. Acoust Soc. Amer. 46, 1549-1555.

Lubker, James F. (1968) An electromyographic-cineradiographic investigation of velar function during normal speech production. Cleft Palate J. 5, 1.

Lubker, James F., Bjorn Fritzell, and J. Lindqvist. (1970) Velopharyngeal function: An electromyographic study. Royal Institute of Technology, STL-QPSR 4, 9-20.

Moll, Kenneth L. (1962) Velopharyngeal closure on vowels. J. Speech and Hearing Res. 5, 30.

Moll, Kenneth L. and Thomas H. Shriner (1967) Preliminary investigation of a new concept of velar activity during speech. Cleft Palate J. 4, 58.

Perkell, Joseph S. (1969) Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. Cambridge, Mass.: M.I.T. Press.

Port, Diane. (1971) The EMG data system. (See this Status Report.)

An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels

Lawrence J. Raphael*
Haskins Laboratories, New Haven

INTRODUCTION

Traditional phonetic literature presents us with a picture of vowel articulation often referred to as the vowel triangle or quadrilateral. In part, this vowel triangle appears as follows for English vowels:

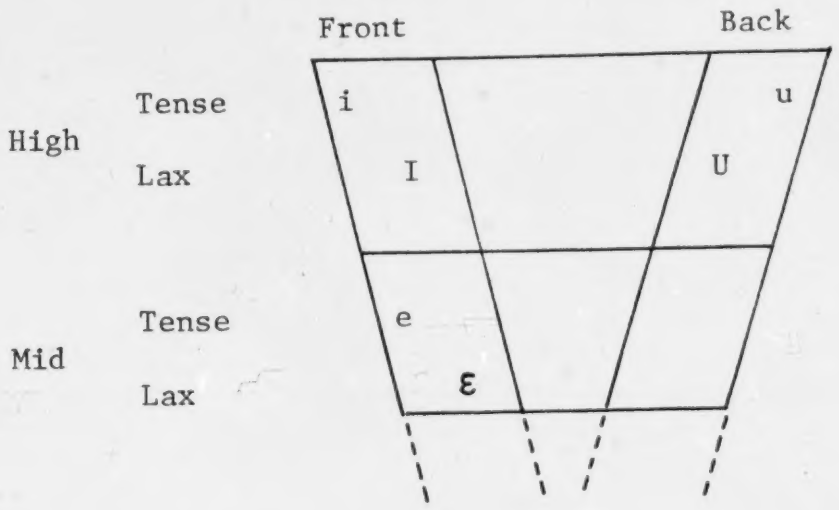


Fig. 1: A Portion of the Vowel Triangle for English Vowels

The two pairs of high vowels and the pair of mid vowels are said to enter into a tense-lax relationship in which the higher member of each pair is articulated with greater muscular effort than the lower member. The difference in height in such a view is often interpreted as a reflection of the difference in tongue tension.[1]

It has been suggested (Perkell, 1969) that the differences in muscular tension between the members of a tense-lax pair of vowels are attributable to the actions of the extrinsic muscles which position the tongue in the oral cavity. The experiment described below was designed to test the traditional

---

*Also Herbert H. Lehman College of the City University of New York.

[1]Other features or combinations of features have been put forth as distinctive in the opposition of these pairs of vowels. Thus they are variously described as long-short (with regard to duration), diphthong-monophthong, close-open (with regard to jaw opening), high-low (independent of the tenseness feature), and free-checked (with regard to their occurrence in English words, without direct reference to their articulation).

tense-lax hypothesis vis-à-vis the vowel triangle, principally in terms of the action of the genioglossus, one of the major extrinsic tongue muscles, to front and bunch the tongue.

PROCEDURE

The corpus of test utterances consisted of the six vowels shown in Figure 1.[2] The vowels were produced in a /əCVC/ context. The initial consonant was /p/ and the final consonants were variously /p,b,t,d,k,g,s,z/. Each vowel was paired with each of the final consonants, yielding a total of forty-eight utterance types. In addition, a small set of twelve utterances was produced, consisting of /ə/, followed by /t/, followed by each of the six vowels, followed by either /p/ or /b/. Thus a total of sixty utterance types was produced in the experiment. The utterances were grouped into two lists of thirty each. Each of the two groups was randomized in several ways. Fifteen tokens of each utterance type were analyzed. The activity of the genioglossus (and of the other muscles considered here) was inferred from the EMG signal transmitted by hooked-wire electrodes inserted into the muscle by means of a hypodermic needle. The insertions are described by Hirose (1971) and the data processing by Port (1971).

RESULTS AND DISCUSSION

The data derived from the action of the genioglossus clearly reflect a tense-lax difference along the traditional lines mentioned above. The differences are most clearly observable in the /əpVp/ and /əpVb/ syllables, since in these cases neither the initial nor the final consonant involves genioglossus activity (Figures 2-4). (The zero point in these figures refers to the onset of voicing of the stressed vowel.)

The data do not, however, arrange the vowels in a manner congruent with a picture of the traditional vowel triangle. A comparison of the peak values of genioglossus activity (Table I) reveals that the front vowels are resolved into two groups: /i-e/ and /I-ɛ/. To whatever extent the genioglossus activity does reflect tongue height, it would appear that the data present a picture in which the vowels are arranged in the following order:

i
e
I
ɛ,

although the differences between /i-e/ and between /I-ɛ/ are often small and occasionally in a direction opposite to that suggested by the ordering given above. In any event, /I/ and /e/ are clearly transposed from their usual positions in the vowel triangle.

---

[2]These vowels were chosen because they are the ones most generally agreed upon as being paired. Although the literature contains claims that various other pairs exist, almost all writers posit those pairs which are investigated here.

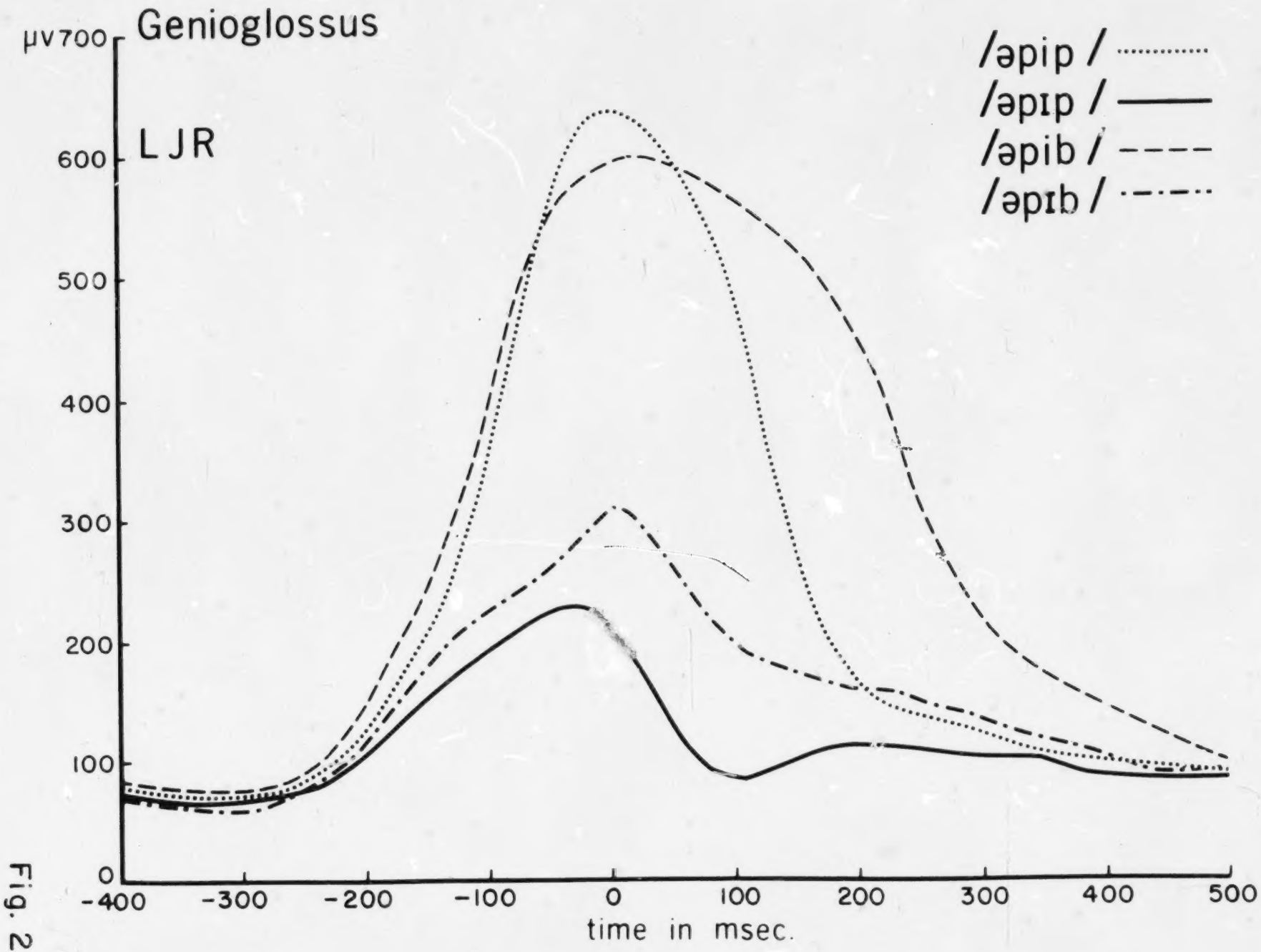

Genioglossus
μv700
600
500
400
300
200
100
0
L J R
/əpip / .....
/əpɪp / ——
/əpib / - - - -
/əpɪb / -·-·-·
-400
-300
-200
-100
0
100
200
300
400
500
time in msec.

Fig. 2

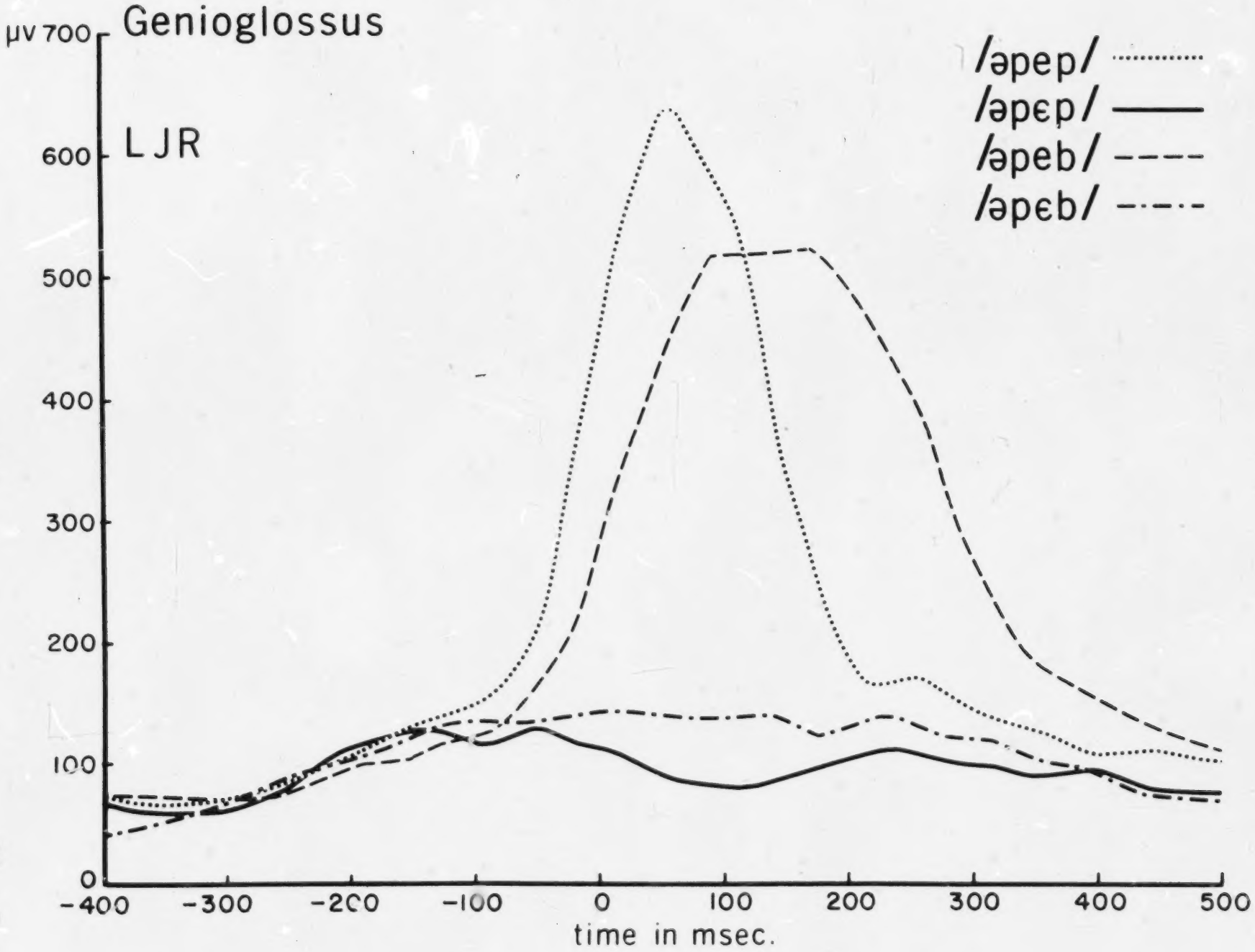

Genioglossus
μv 700
600
500
400
300
200
100
0
LJR
/əpep/
/əpɛp/
/əpeb/
/əpɛb/
-400
-300
-200
-100
0
100
200
300
400
500
time in msec.

Fig. 3

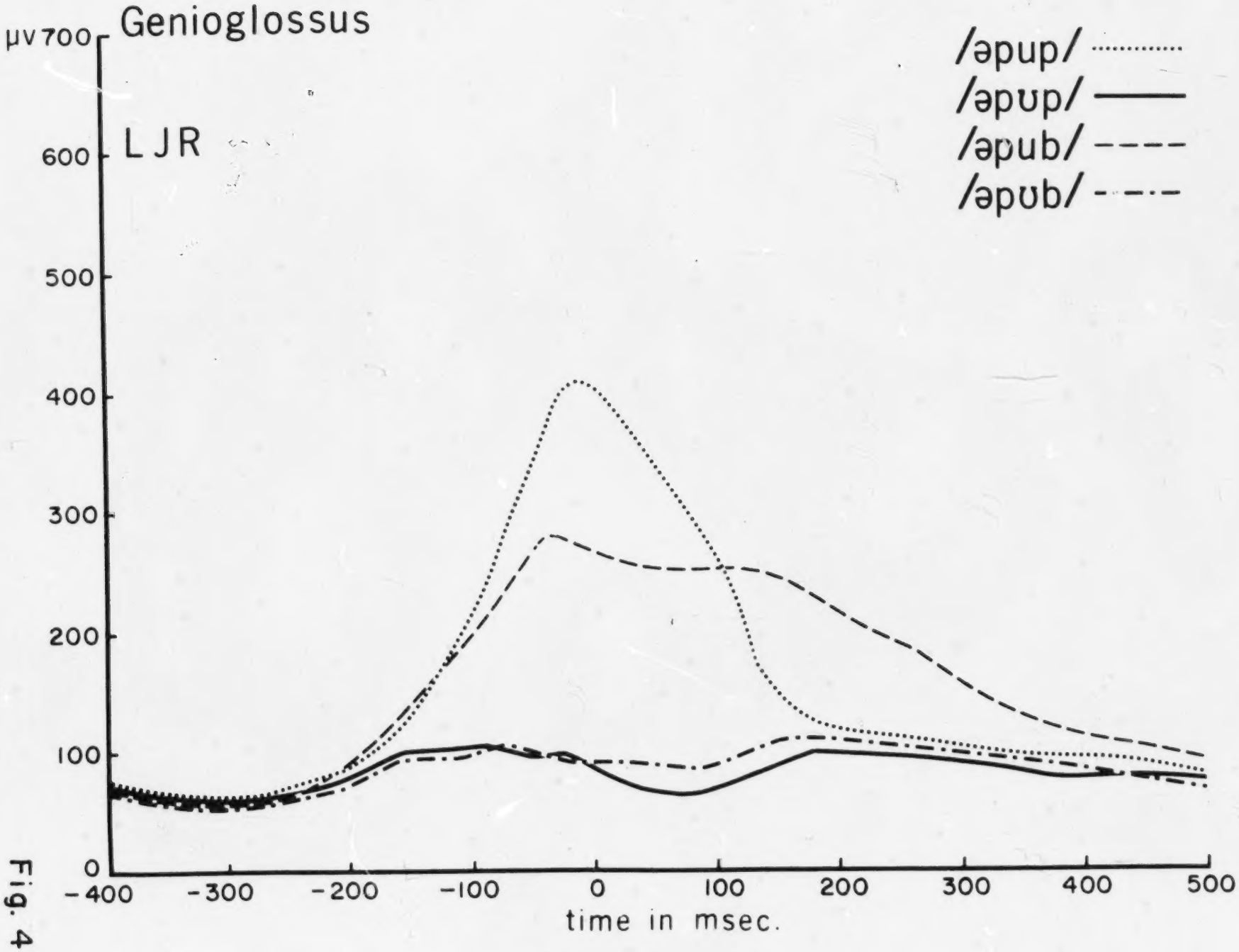

Genioglossus
μv 700
600
500
400
300
200
100
0
L J R
/əpup/
/əpʊp/
/əpub/
/əpʊb/
-400
-300
-200
-100
0
100
200
300
400
500
time in msec.

Fig. 4

TABLE I

PEAK VALUES IN MICROVOLTS OF THE EMG SIGNAL
FROM THE GENIOGLOSSUS DURING
VOWEL ARTICULATION

VOWEL	CONSONANT CONTEXT									
	p-p	p-b	p-t	p-d	p-k	p-g	p-s	p-z	t-p	t-b
i	653	606	702	644	694	625	532	537	582	579
I	236	319	316	360	268	494	223	294	271	243
e	643	533	710*	538*	643	571	530	500	562	447
ɛ	140	154	179*	331*	206	481	172	195	108	153
u	410	289	433	342	302	257	325	246	380	302
U	112	110	143	251	60	100	118	148	68	85

*These four figures are typical of an effect found generally throughout the data: the tense vowels show a decrease of activity from a following voiceless to a following voiced context; the lax vowels show an increase in activity in the same context. As yet, this effect awaits explanation in the light of further analysis of the data and of additional data from other subjects.

There are at least two possible explanations for this result. The first is based on the assumption that the genioglossus data do not present a complete picture of tongue height. Rather, tongue height is most likely the result of the combination of two factors: (1) tongue bunching, accounted for largely by the activity of the genioglossus,[3] and (2) jaw opening (Lindblom and Sundberg, 1969). That is, a given tongue height, measured from the palate to the high point of the tongue can be attained in more than one way: e.g., wide jaw opening with extreme tongue bunching or narrow jaw opening with minimal tongue bunching. If, in fact, /I/ is a high vowel to be paired with /i/, and if it is higher than /e/, we could expect to find less jaw opening for /I/ than for /e/ to compensate at least partially for the greater tongue bunching of the latter vowel.

[3]That is, in this experiment, since the superior longitudinal is generally recognized as playing a prominent role in this function (MacNeilage and deClerk, 1969).

Among the muscles investigated in this experiment was the sternohyoid, which is described as a muscle accompanying jaw opening (Ohala and Hirose, 1970). The data for the activity of the sternohyoid do consistently reveal a greater jaw opening for /e/ (and /ɛ/) than for /I/ (and /i/) (Table II). In fact, the data for these front vowels generally (but not with complete consistency) show just what the traditional vowel triangle would lead us to expect: increasing values for the series /i,I,e,ɛ/, which we take here to mean increased jaw opening for the vowels as they descend from high to low. Figures 5 and 6 display the data for the labial bounded syllables. The relevant portions of the displays are found between the vertical lines.

TABLE II

PEAK VALUES IN MICROVOLTS OF THE EMG SIGNAL
FROM THE STERNOHYOID DURING
VOWEL ARTICULATION

VOWEL	CONSONANT CONTEXT*							
	p-p	p-b	p-t	p-d	p-k	p-g	p-s	p-z
i	41	44	35	55	62	67	62	52
I	43	53	45	48	58	56	58	57
e	82	74	55	68	96	89	72	96
ɛ	99	91	76	75	113	119	106	105

*Because of the involvement of the sternohyoid in the articulation of /t/, no separate peaks of activity are discernible for the vowels in the /t-p/ and /t-b/ contexts. Thus, they have been omitted from the table.

The second possible explanation for the transposition of /I/ and /e/ in the usual height ordering involves the matter of tongue backing. The vowel triangle (Figure 1) shows both /I/ and /ɛ/ to be retracted from the more extreme front positions of /i/ and /e/. Since the genioglossus displays greater activity the more fronted the tongue is (Hirano and Smith, 1967; also compare the values for /i-e/ vs. those for /u-U/ in Table I above), one would naturally expect lower values for the activity of this muscle for /I/ (and /ɛ/) if, in fact, these vowels are less fronted than are /i/ and /e/.

A muscle tapped in this experiment which is taken to be an indicator of tongue backing is the superior constrictor. The data from this muscle do often reveal a greater degree of tongue retraction for /I/ and /ɛ/ as

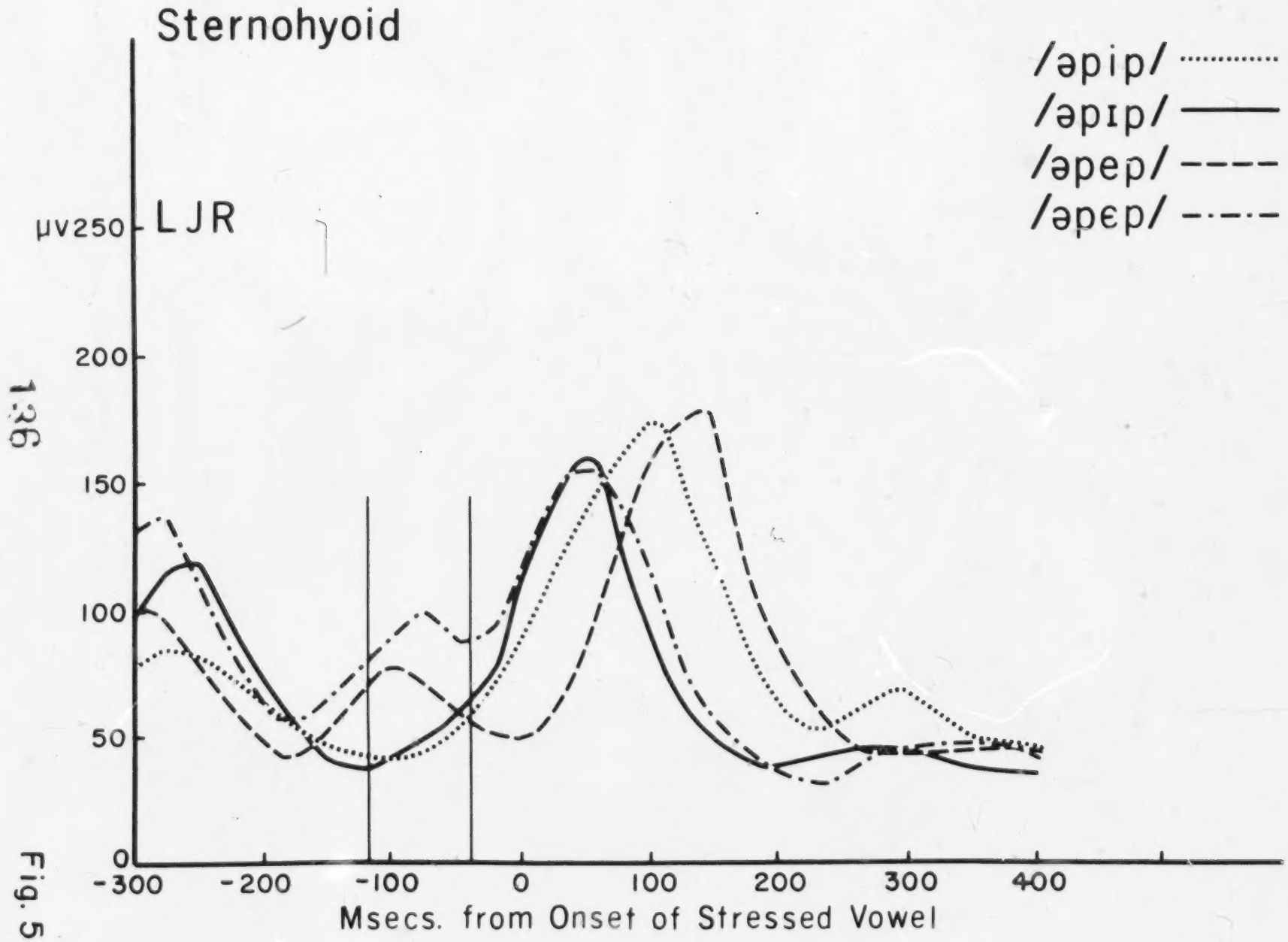

Sternohyoid
/əpip/
/əpɪp/
/əpep/
/əpɛp/
μv250
200
150
100
50
0
LJR
-300
-200
-100
0
100
200
300
400
Msecs. from Onset of Stressed Vowel

Fig. 5 1:36

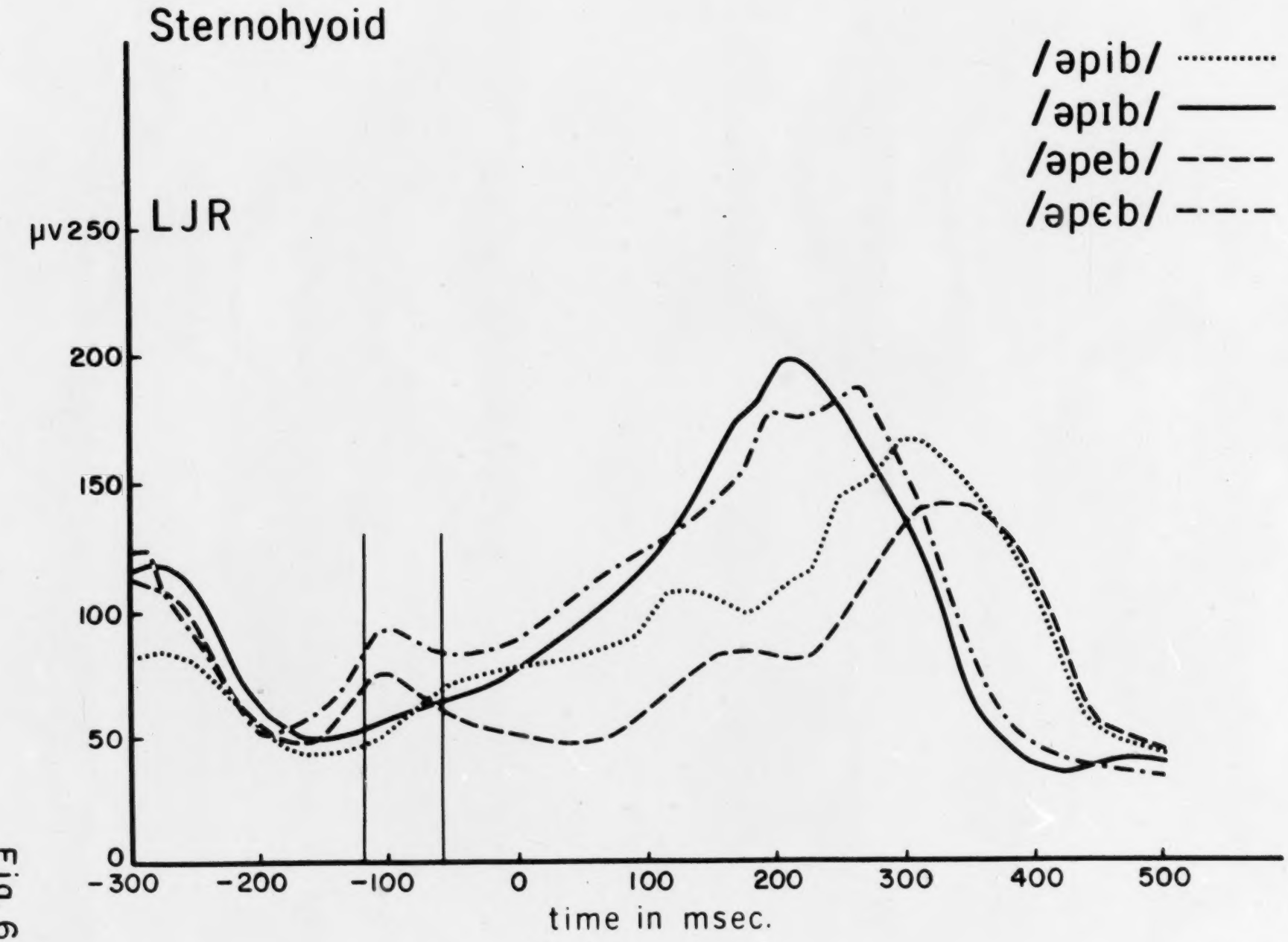

Sternohyoid
μv250
LJR
200
150
100
50
0
-300
-200
-100
0
100
200
300
400
500
time in msec.
/əpib/
/əpɪb/
/əpeb/
/əpɛb/

Fig. 6

1:37

opposed to /i/ and /e/, but the results are not consistent, differences occasionally being small and/or in the unhypothesized direction.

CONCLUSION

The possibilities discussed above, then, reduce but do not eliminate the discrepancy between the usual height ordering of the front vowels and their grouping into tense-lax pairs on the one hand and the data from the experiment for the genioglossus on the other. Although the data do not allow for a strong reaffirmation of the traditional view of the vowel triangle along tense-lax and high-low lines simultaneously, there is some reason to believe that with the addition of more data from other muscles and other subjects, and perhaps with the consideration of other factors besides jaw opening and tongue fronting vs. backing, the traditional picture of vowel articulation may be confirmed.

REFERENCES

Hirano, Minoru and Timothy Smith. (1967) Electromyographic study of tongue function in speech: A preliminary report. Working Papers in Phonetics 7 (Los Angeles: UCLA Phonetics Laboratory) 45-56.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. (See this Status Report.)

Lindblom, Bjorn and J. Sundberg. (1969) A quantitative model of vowel production and the distinctive features of Swedish vowels. Quarterly Progress and Status Report 1 (Stockholm: Dept. of Speech Communication, Speech Transmission Laboratory, Royal Institute of Technology) 14-32.

MacNeilage, Peter F. and Joseph L. deClerk. (1969) On the motor control of coarticulation in CVC monosyllables. J. Acoust. Soc. Amer. 45, 1217-1233.

Ohala, John and Hajime Hirose. (1970) The function of the sternohyoid muscle in speech. Annual Bulletin, Research Institute of Logopedics and Phoniatrics 4 (Tokyo: University of Tokyo) 41-44.

Perkell, Joseph S. (1969) Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. Cambridge, Mass: M.I.T. Press.

Port, D. (1971) The EMG data system. (See this Status Report.)

Vocal Tract Size Normalization in the Perception of Stop Consonants*

Timothy C. Rand+
Haskins Laboratories, New Haven

One type of acoustic variation confronting a listener stems from the fact that vocal tract dimensions vary from speaker to speaker. Speakers with smaller vocal tracts generally produce speech with higher formant frequencies. This is apparent in Peterson and Barney's (1952) formant frequency data for vowels produced by men, women, and children.

Ladefoged and Broadbent (1957) have demonstrated context effects on vowel perception, where certain of the variations in the contextual material reflected variations in vocal tract size. The listeners interpreted vowels and preceding carrier phrases as if they had been produced by the same sized vocal tract. These results establish the existence of vocal tract normalizing functions in the human speech perception mechanism. Additional support is provided by Darwin's (1971) demonstration of a right-ear effect when vowels are presented dichotically and vocal tract size is varied from trial to trial.

Consonant perception differs markedly from vowel perception, particularly in the case of the stop consonants. A stop displays a significant lack of acoustic invariance when it occurs together with different vowels. For the place of articulation dimension, this context-conditioned variation has been explained with reference to acoustic loci to which the second formant transition "points" (Delattre et al., 1955). The loci, which are not directly realized in the acoustic speech signal, can be considered to correspond to the resonant frequencies of the occluded vocal tract. Stevens and House (1956) used a vocal tract analog synthesizer to test this hypothesis by measuring resonances when the tract was appropriately constricted. Their findings for second formant loci agree well with the experimental results of the Haskins group.

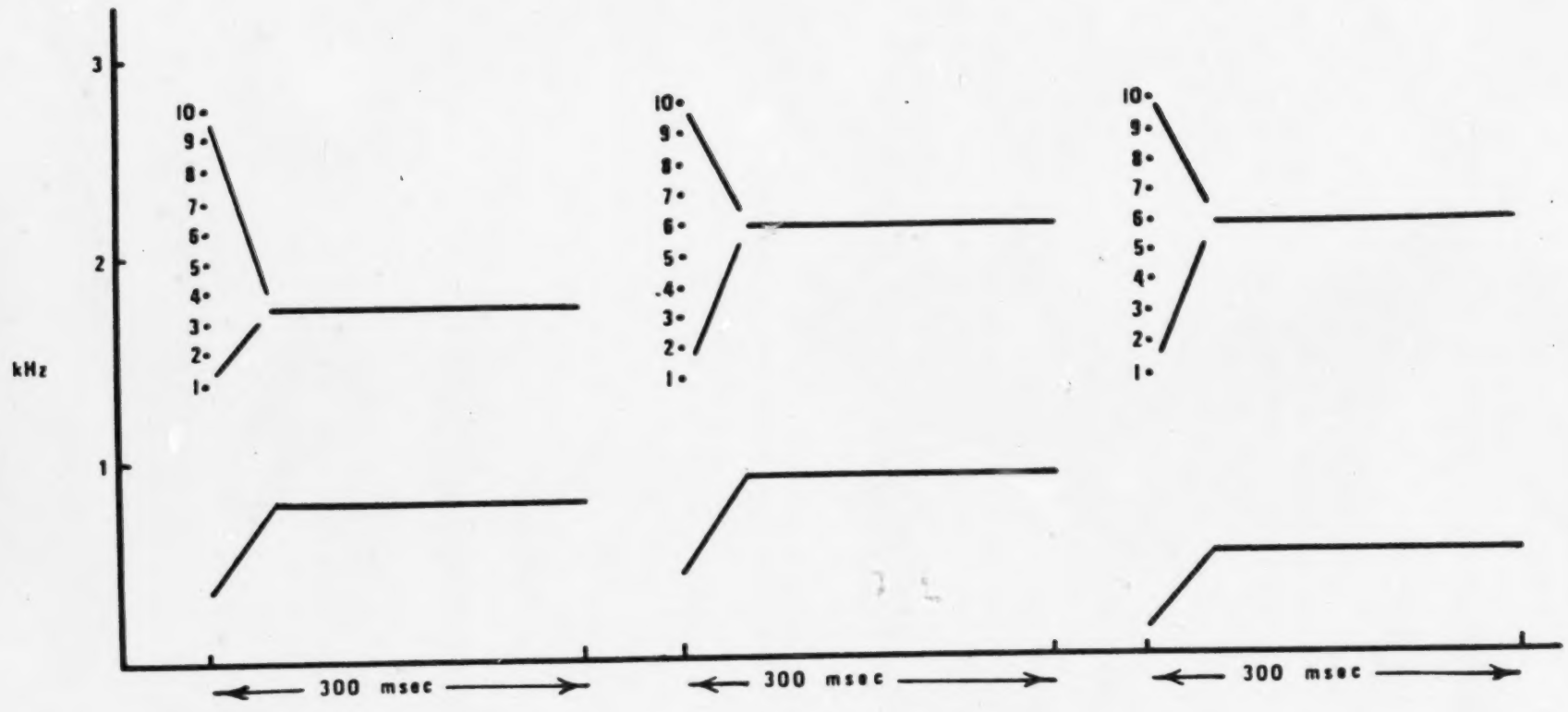
If the locus concept is to be related to the resonance of the occluded vocal tract, then it is to be expected that this resonant frequency will vary with vocal tract size. Fourcin (1968) demonstrated that prior context influences the perception of synthetic whispered consonant-vowel syllables. Fourcin interpreted his results as a shift in consonantal locus depending upon whether a precursive "hallo" was spoken by a man or a child.

---

*Paper presented at the 81st meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

+Also the University of Connecticut, Storrs.

142


3
2
1
kHz
10
9
8
7
6
5
4
3
2
1
300 msec
300 msec
300 msec
LARGE VOCAL TRACT
[c æ]
SMALL VOCAL TRACT
[c æ]
LARGE VOCAL TRACT
[c ɛ]

140

Fig. 1

A further test of locus shift would be to see whether the second formant transitions of CV syllables can cue different consonants depending upon vocal tract information carried on the syllable itself.

The stimuli used in the present experiment (Fig. 1) were thirty two-formant synthetic CV syllables. All syllables were of 300-msec duration and were produced on the Haskins parallel formant synthesizer. The stimuli fall into three groups, based upon the characteristics of the vocalic portion of the syllable. In the first group, the steady-state vowel formants are appropriate to [æ] as rendered by an adult speaker. The vowel formants of the second group of stimuli are related to those of the first group by a multiplicative constant (approximately 1.2); these are roughly appropriate to a child's rendition of [æ]. The vowel of the third stimulus group is roughly [ɛ] as rendered by an adult. Groups 2 and 3 are related in that they have identical second formants.

All stimuli have initial rising first formant transitions to serve as a cue for voiced manner of production. As regards second formant transitions, ten equally spaced frequencies were chosen, ranging from 1386 to 2762 Hz, to serve as initial transition frequencies.

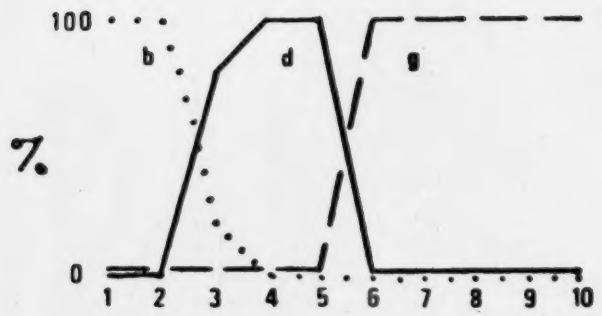
A tape was prepared with five repetitions of each stimulus (150 items total) in a random sequence. Listeners were told they would hear synthetic CV syllables and were instructed to write down the initial consonant for each.

Figure 2 displays identification functions for two subjects. The functions are plotted separately for the three classes of stimuli. In each case, the ordinate is percent identification and the abscissa is the stimulus number, indicating the starting point of the second formant transition. There is a clear tendency for the consonant categories to occur at higher frequencies for the small vocal tract stimuli than for both varieties of large vocal tract stimuli.

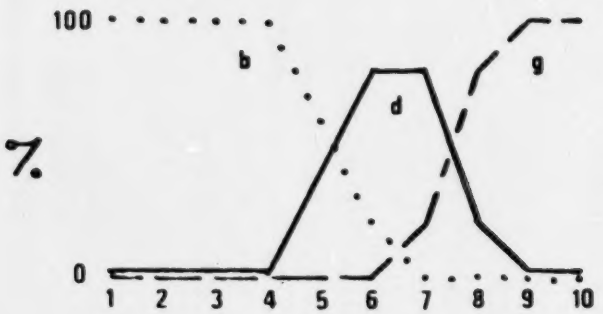
Another way to observe how vocal tract size affects place identification is to compute the mean [d] response on the ten-point scale corresponding to initial second formant transition frequency. For example, for the subject whose data are displayed on the left in Figure 2, the mean [d] responses are 4.1, 6.4, and 4.0 for the three stimulus groups going from top to bottom. These means were computed for ten subjects and are displayed in Figure 3.

In Figure 3, the "S's" indicate the mean [d] response for each subject for the Small vocal tract stimuli; similarly, the "L's" indicate the mean values for the two conditions involving Large vocal tract stimuli. The important feature of these results is that, for all ten subjects, small vocal tract stimuli produced the response [d] for higher values of second formant transition frequency than did the large vocal tract stimuli.

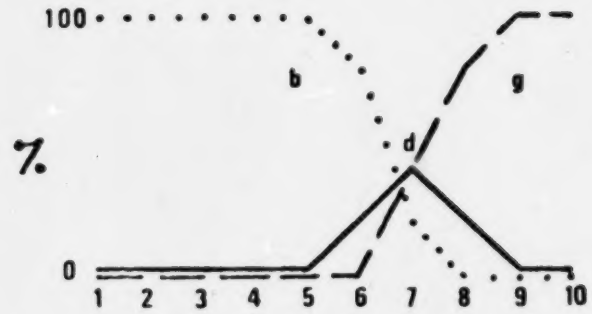
Speech utterances are effective communicators whether they originate from small children or large adults. Normalization for vocal tract size differences during the process of speech perception amounts to a cancellation of one type of acoustic variance that is introduced during the production phase of the speech chain. One way to characterize this ability on the part of a listener is to say that he somehow extracts or reconstructs the speaker's

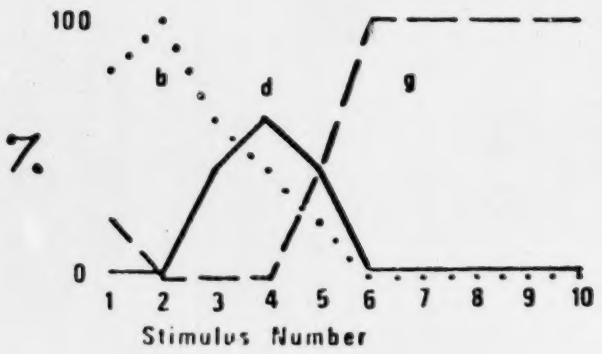


LARGE
VOCAL TRACT
[Cæ]

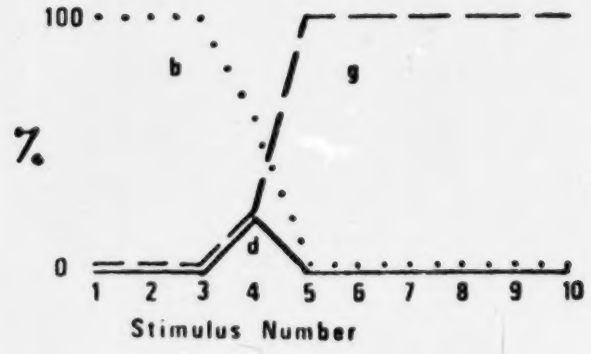


SMALL
VOCAL TRACT
[Cæ]





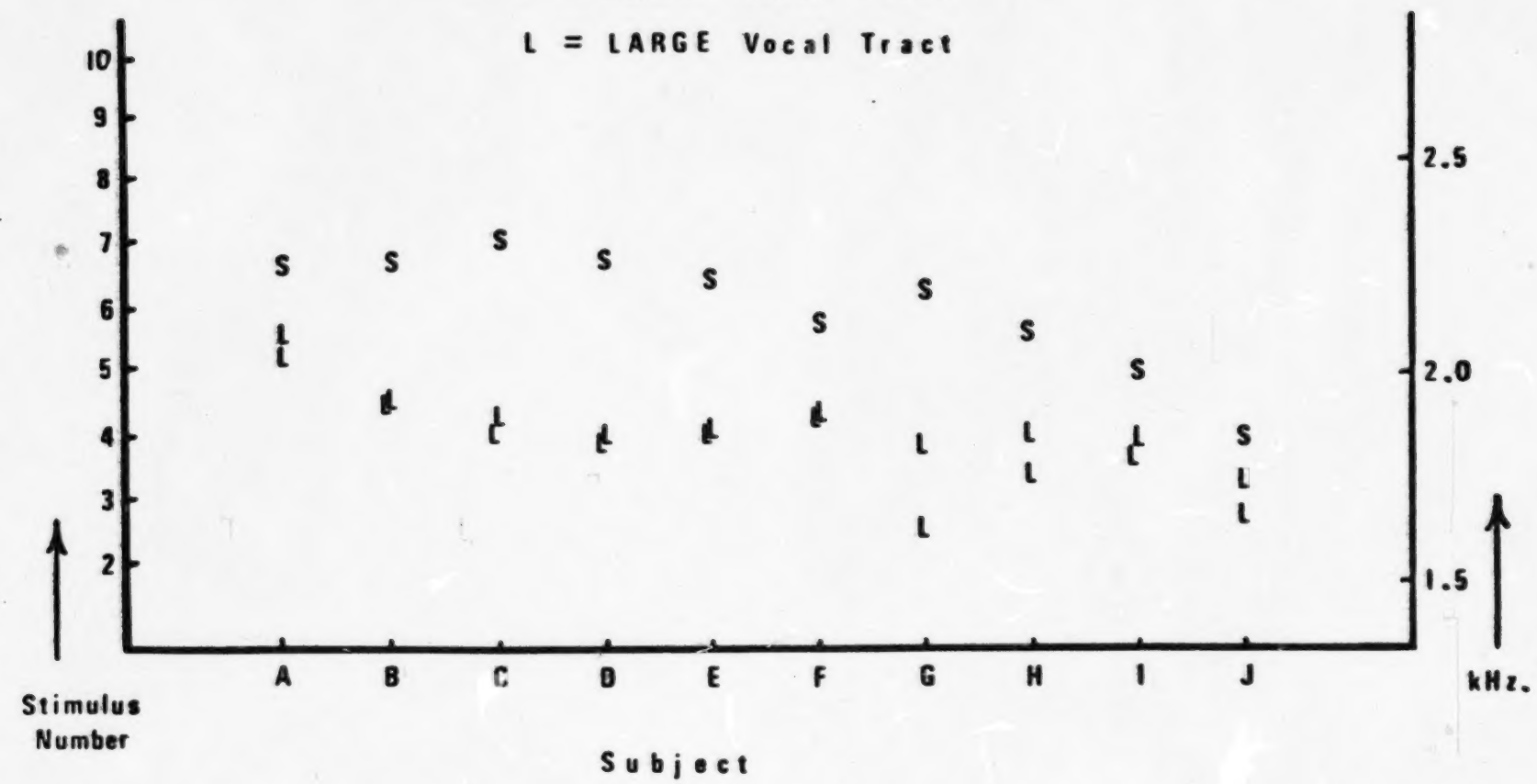
LARGE
VOCAL TRACT
[Cɛ]



IDENTIFICATION

142

Fig. 2



MEAN [d] RESPONSE

Fig. 3

articulatory intent. A motor theory of speech perception, such as that advanced by the Haskins workers, is based on the observation that perception is often more closely related to articulation than to the acoustic signal. This view provides a framework within which the results of the present study receive a natural interpretation. If it can be said that listeners perceive consonants by reference to locus frequencies, then the subjects of this experiment perceived consonants by reference to loci appropriate to the vocal tract sizes producing the syllables they heard.

REFERENCES

Darwin, C. (1971) Ear differences in the recall of fricatives and vowels. Quart. J. of Exp. Psychol. 23, 46-62.

Delattre, P.C., A.M. Liberman, and F.S. Cooper. (1955) Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer. 27, 769-773. (Also in Lehiste, 1967).

Fourcin, A.J. (1968) Speech source inference. IEEE Transactions AU-16, 65-67.

Ladefoged, P., and D.E. Broadbent (1957) Information conveyed by vowels. J. Acoust. Soc. Amer. 29, 98-104. (Also in Lehiste, 1967).

Lehiste, I. (ed.). (1967) Readings in Acoustic Phonetics. Cambridge: MIT Press.

Peterson, G.E., and H.L. Barney. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Amer. 24, 175-184. (Also in Lehiste, 1967).

Stevens, K.N., and A.S. House. (1956) Studies of formant transitions using a vocal tract analog. J. Acoust. Soc. Amer. 28, 578-585.

ABSTRACT

Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English*

Lawrence Raphael+
Haskins Laboratories, New Haven

Much research by linguistis and phoneticians has been directed toward discovering a single physiological or acoustic basis which underlies the perceptual separation of cognate sounds. Experimental evidence indicates that neither of the two most commonly suggested bases, the voiced/voiceless opposition and the fortis/lenis opposition, has enough generality of distribution to be a unique explanation of cognate perception. There is, however, a substantial number of different acoustic and articulatory cues which can account for cognate perception in virtually all phonetic environments. These cues, some of which are thought of as reflexes of the voiced/voiceless or fortis/lenis oppositions, include stop release, aspiration, stop closure duration, friction duration, intraoral air pressure, preceding vowel duration, presence or absence of fundamental frequency, and timing relationships between articulation and glottal states. The last of these cues, currently expressed as voicing onset time (VOT) in the case of prevocalic stops, may be sufficiently extended in its generality to account for cognate perception in all phonetic environments. Or, there may be no way to generally account for the phenomenon.

The relationship between articulation and glottal states has received little attention for sounds in absolute final position. It is possible that voicing offset time, the mirror-image analogue of VOT, is a significant cue to cognate perception in this position. Spectrographic measurements of voicing offset time reliably separate cognate categories of stops and fricatives. Tape-cutting experiments, however, in which subjects heard a randomized series of real-speech words from which varying degrees of voicing during final /bdg/ closure was removed, revealed no significant changes in perception in the direction of /ptk/, even when the closure period was completely eliminated and a small portion of the preceding vowel transition was cut away.

Stop and fricative cognates are also reliably separated by spectrographic measurements of preceding vowel duration. Tape-cutting experiments on real speech, in which the vowels before final stops and fricatives were shortened, caused perception to change from /bdgvðzʒ/ to /ptkfθsʃ/. Secondary perceptual reversals, for stimuli in which the vowel was virtually eliminated,

---

*Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

+Also City University of New York.

indicated that voicing during final consonant closure does have minor cue value.

A variety of minimal and subminimal CVC(C) pairs was synthesized on the Haskins Laboratories' Pattern Playback. The vowel duration in each was varied over a range of values derived from those found in real speech. It was found that, regardless of the cues for voicing or voicelessness used in the synthesis of the final consonant or cluster, listeners perceived the final segments as voiceless when they were preceded by vowels of shorter duration and as voiced when they were preceded by vowels of longer duration. Discrimination tests revealed that perception across and within phoneme boundaries was continuous rather than categorical. It was also found that the cue of vowel duration is more effective before final stops and clusters than before final fricatives. The indication that voicing during final consonant closure does have minor cue value for cognate perception received further confirmation in the synthetic-speech experiment.

It is concluded that preceding vowel duration is both a sufficient and necessary cue to the perception of the voicing characteristic of word-final consonants and clusters: that is, word-final cognate sounds are perceptually separated and identified on the basis of the duration of the vowels which precede them. Such a finding runs counter to the traditional assumption that features which identify linguistic segments are produced within the articulatory period of those segments. Thus, traditionally, the expectation would have been that the perception of final voiced stops and clusters, for example, results from the listener's recognition of vocal pulsing during consonant closure. Instead, this study demonstrates that such a perception depends primarily upon the listener's recognition of a suitably long vowel preceding the closure for the final consonant or cluster.

PUBLICATIONS AND REPORTS*

Publications and Manuscripts

Ear Differences in the Recall of Fricatives and Vowels. Christopher J. Darwin. Quarterly Journal of Experimental Psychology, 23, 46-62.

Some Recent Developments in the Use of Electromyography in Speech Research. Thomas Gay and Katherine S. Harris. Journal of Speech and Hearing Research, 14, 241-246.

Auditory Evoked Potentials During Speech Perception. Charles C. Wood, William R. Goff, and Ruth S. Day. To appear in Science.

Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability. Isabelle Y. Liberman. To appear in Annual Bulletin of the Orton Society.

An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage. C. J. Darwin, M. T. Turvey, and R. G. Crowder. To appear in Cognitive Psychology.

Electromyography of the Intrinsic Laryngeal Muscles During Phonation. Thomas Gay, Hajime Hirose, Marshall Strome, and Masayuki Sawashima. Submitted for publication.

Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening. C. J. Darwin.

Reports and Oral Papers

Colloquium. Alvin M. Liberman. New York University Medical School. January 1971.

Speech Acoustics, Perception and the Evolution of Language. Philip Lieberman. Brown University, Providence, 13 January 1971.

What Constitutes Perceptual Competition in Dichotic Listening? Ruth S. Day and James E. Cutting. Paper presented at the 42nd Annual Meeting of the Eastern Psychological Association, New York, 15-17 April 1971.

Neural Specialization for Speech Perception. Michael Studdert-Kennedy. Colloquium, University of Texas, Austin, 19 April 1971.

Perception of Dichotically Presented Stop Consonants as a Function of Kind and Amount of Information Shared by Ears. Donald Shankweiler, Michael Studdert-Kennedy, and David Pisoni. Oral Paper presented at the 81st Meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

---

*Most of the contents of this report, SR 25/26, are included in this listing.

Dichotic Forward and Backward Masking of Complex Sounds, C. J. Darwin. Oral paper presented at the 81st Meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables. Robert J. Porter. Paper presented at the 81st Meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

Vocal Tract Size Normalization in the Perception of Stop Consonants. T. C. Rand. Paper presented at the 81st Meeting of the Acoustical Society of America, Washington, D.C., 20-23 April 1971.

Cross-Language Testing of Distinctive Features in Production and Perception. Arthur S. Abramson. Talk at the University of Rochester, 30 April 1971.

What Can Speech Scientists and Speech Pathologists Learn from Each Other? Michael Studdert-Kennedy. Panel Discussion, New York Speech and Hearing Association, Monticello, 4 May 1971.

How is Language Conveyed by Speech? Franklin S. Cooper. Paper presented at the NICHD Growth and Development Program Conference, Communicating by Language: The Relationships Between Speech and Learning to Read, Elkridge, Md., 16-19 May 1971.

How is Reading Grafted onto a Language that the Reader Has Already Acquired? Ignatius G. Mattingly. Paper presented at the NICHD Growth and Development Program Conference, Communicating by Language: The Relationships Between Speech and Learning to Read, Elkridge, Md., 16-19 May 1971.

A Comparison of the Way Children Acquire Spoken and Written Language. Donald Shankweiler and Isabelle Liberman. Paper presented at the NICHD Growth and Development Program Conference, Communicating by Language: The Relationships Between Speech and Learning to Read, Elkridge, Md., 16-19 May 1971.

The EMG Data System. Diane K. Port.

Electromyography of the Articulatory Muscles: Current Instrumentation and Technique. Hajime Hirose.

Action of the Extrinsic Musculature in the Control of Tongue Position: Preliminary Report. Katherine S. Harris.

Electromyography of the Intrinsic Laryngeal Muscles During Phonation. Thomas Gay, Hajime Hirose, Marshall Strome, and Mayasima Sawashima.

An Electromyographic Study of Laryngeal Adjustments During Speech Articulation: A Preliminary Report. Hajime Hirose.

The Velopharyngeal Mechanism: An Electromyographic Study -- A Preliminary Report. Fredericka B. Berti.

An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels. Lawrence J. Raphael.

Thesis

Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English. Lawrence J. Raphael. Ph.D. dissertation, City University of New York.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Haskins Laboratories, Inc.
270 Crown Street
New Haven, Conn. 06510

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

N/A

3. REPORT TITLE

Status Report on Speech Research, No. 25/26, January-June 1971

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories; Franklin S. Cooper, P.I.

6. REPORT DATE

July 1971

7a. TOTAL NO. OF PAGES

147

7b. NO. OF REFS

219

8a. CONTRACT OR GRANT NO.

ONR Contract N00014-67-A-0129-0001

b. PROJECT NO.

NIDR: Grant DE-01774
NICHD: Grant HD-01994

c. NIH/DRFR: Grant FR-5596
VA/PSAS Contract V-1005M-1253

d. Conn. Res. Com.: Grant RSA-70-9

9a. ORIGINATOR'S REPORT NUMBER(S)

SR-25/26 (1971)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (for 1 January-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts and extended reports cover the following topics:

-Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study
-An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage
-Stimulus Versus Response Competition and Ear Asymmetry in Dichotic Listening
-The Effect of Temporal Overlap on the Perception of Dichotically and Monotically Presented CV Syllables
-Auditory Evoked Potentials During Speech Perception
-Basic Research in Speech and Lateralization of Language: Some Implications for Reading Disability
-The EMG Data System
-Electromyography of the Articulatory Muscles: Current Instrumentation and Technique
-Action of the Extrinsic Musculature in the Control of Tongue Position
-Electromyography of the Intrinsic Laryngeal Muscles During Phonation
-An Electromyographic Study of Laryngeal Adjustments During Speech Articulation
-The Velopharyngeal Mechanism: An Electromyographic Study
-An Electromyographic Investigation of the Tense-Lax Feature in Some English Vowels
-Vocal Tract Size Normalization in the Perception of Stop Consonants
-Vowel Duration as a Cue to the Perceptual Separation of Cognate Sounds in American English

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Dichotic listening experiments						
Speech and nonspeech processing						
Speech perception						
Auditory storage						
Lateralization of language						
Reading disabilities						
Speech physiology						
EMG instrumentation and techniques						
EMG in speech articulation						